Financial Accountability and Managerial Incentives in English NHS Hospital Trusts 2003-2008

Greenwood, Margaret

Award date:
2012

Awarding institution:
University of Bath

FINANCIAL ACCOUNTABILITY AND MANAGERIAL INCENTIVES IN ENGLISH NHS HOSPITAL TRUSTS 2003-2008

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

MARGARET JOAN GREENWOOD MA, BSC, FCA

UNIVERSITY OF BATH SCHOOL OF MANAGEMENT

FEBRUARY 2012

FINANCIAL ACCOUNTABILITY AND MANAGERIAL INCENTIVES IN ENGLISH NHS HOSPITAL TRUSTS 2003-2008

ABSTRACT

A continuing programme of NPM reforms, grounded in quasi-market modes of governance and private sector best practice, have been applied to English NHS hospitals over the last thirty years in response to concerns about their performance efficiency and accountability. However, in the transition to market modes of governance, the retention of hierarchical features gave rise to a multi-layering of accountability. From 2001-02 balanced scorecard inspired performance measurement systems (PMS), were introduced into the NHS, aimed at improving service standards through improved cost efficiency. Study 1 in this thesis finds that, in this context, the relationship between service standards and cost efficiency is positive and that, consistent with it being a more effective PMS, this was stronger for the 'Annual Health Check', a PMS characterised by features aimed at reducing manipulation, than the Star ratings, its predecessor.

The approach to the manipulation of financial breakeven, a key accountability measure, was however more relaxed, particularly when service standards were under threat. The system of 'financial support' had its roots in previous hierarchical relationships and acted to shift revenue across the NHS in order to allow Trusts in financial difficulty to meet their financial objectives without damaging service standards. These transfers, which were effected through the revenue account, were generally reversed out in future years with the result that financial support accelerated revenue recognition in Trusts receiving it. In Study 2, the receipt of financial support by Trusts in financial difficulty was found to be associated with an improvement in service standards and in future financial performance but, in an increasingly demanding performance regime and multi-layered accountability, evidence was also found of opportunistic exploitation of the system. Financial support had a considerable impact on the accountability of both NHS Trusts and the wider NHS because of the limited transparency around financial support transactions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF CASE STUDIES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AGM	Annual General Meeting
AHC	Annual Health Check
BSC	Balanced Scorecard
CEO	Chief Executive Officer
DEA	Data Envelopment Analysis
ENT	Existing National Targets
GAAP	Generally Accepted Accounting Practice
NAO	National Audit Office
NHS	National Health Service
OLS	Ordinary Least Squares
NPM	New Public Management
PCT	Primary Care Trust
PFI	Private Finance Initiative
PMS	Performance Measurement System
SFA	Stochastic Frontier Analysis
SHA	Strategic Health Authority
SLA	Service Level Agreement
UoR	Use of Resources

DEDICATION

For my husband, to whom, miraculously, I am still married.

ACKNOWLEDGEMENTS

I would first like to thank my supervisor, Professor John Forker, for his continuing support, advice and guidance during the period of my research apprenticeship. Through his encouragement and inspiration I have been able to discover both the joys and the challenge of research.

I would also like to acknowledge the helpful comments received from my colleagues at the University of Bath, from seminar participants at the University of Bangor in 2011, at the University of Bristol in 2010, at Queen's University, Belfast, in 2007, and from participants at the American Accounting Association's Annual Meeting in San Francisco in 2010 and at the European Accounting Association's Annual Congress in Istanbul, also in 2010. Finally, I would like to thank Professor S. Young for his comments on the earnings management aspects of Study 2.

CHAPTER 1 INTRODUCTION

1.1 PERSONAL MOTIVATION

Having previously had an accounting career in the private sector I spent ten years, from 1997 to 2006, as a non-executive director on the Board of a number of NHS Trusts, including three years in the role of Chairman. This period embraced the a number of significant reforms to the healthcare system introduced by the Labour Government from the early 2000s and I observed the incentive impact of these reforms, particularly those related to performance measurement and management, on the Trust and its executive managers.

During my period as a Board member, the management of the Trust felt very corporate, particularly in relation to its governance structures which emulated the private sector unitary board model. However, although we 'felt' like self -governing institutions contracting with healthcare commissioners, there were aspects of accountability relationships which intrigued me. Why, for example, when we held a private sector style AGM and were publicly accountable in our own right, were the Chair and CEO annually summonsed to the offices of the Department of Health's Regional Office and required to answer to the Regional Director? And in relation to financial accountability and the financial statements, why were the auditors so relaxed about accounting policies that seemed to be at odds with generally accepted accounting practice?

This thesis was inspired by these questions and extends previous research which was focused largely on governance and on the manipulation of financial performance (Ballantine, Forker and Greenwood, 2007, 2008a and 2008b). The main focus of the thesis is on the incentive impact of performance measurement and on the effectiveness of permitting flexibility in reporting performance, both financial and nonfinancial[1].

[1] Although inspired by my experience on the Boards of NHS Trusts, the specific Trusts with which I have been associated have not contributed in any way to this thesis, nor are they associated with the analysis or conclusions, which are entirely my own.

1.2 PRIMARY THEME

A large and diverse literature, based in agency theory, explores the question of how best to incentivise managerial effort in the interests of principals. Recognition that over-reliance on financial measures and their potential for being manipulated in ways which can undermine the strategic capacity of the organisation (Hopwood, 1972) has led to the development of multi-dimensional models of performance measurement such as the balanced scorecard (Kaplan and Norton, 1992, 1996a). These models, by introducing nonfinancial measures of performance, aim to incentivise goal congruent managerial effort and to mitigate managerial opportunism in the achievement of financial accountability objectives. A key assumption underpinning these models is that additional measurement diversity is beneficial, an assumption that is supported by analytical studies (Holmstrom, 1979; Banker and Datar, 1989; and Feltham and Xie, 1994). However, in both the analytical and normative literature a proliferation of performance measures is seen as counterproductive, diverting and dispersing management effort away from key strategic objectives (Kaplan and Norton, 1992) and potentially generating no benefits or incurring unnecessary costs (Feltham and Xie, 1994).

The manipulation of performance measures is the subject of an extensive empirical literature which, in a variety of institutional settings, seeks to identify such manipulation and portray it as having adverse consequences. In the NHS Bevan and Hood (2006) provide prima facie evidence of manipulation to achieve performance targets in the NHS and argue that this renders these measures unreliable for the purposes of assessing performance and the delivery of public accountability. Elsewhere, however, studies have found that the focus on waiting targets and the improvements observed in the English NHS have not been achieved at the expense of important unmeasured activities (Propper, Sutton, Whitnall and Windjmeier, 2010; Kelman and Friedman, 2009) but that, on the contrary, other aspects of performance, such as mortality and readmission rates, not integrated into the performance measurement system, also improved.

The issue of whether manipulation of reported performance can be beneficial and in the principal's interest has been more fully explored in the accounting literature where a more balanced perspective of manipulation is adopted. This literature argues, and demonstrates, that, even though it undermines

financial accountability, the manipulation of reported financial performance can be in the interests of the principal.. Demski, Frimor and Sappington (2004) for example demonstrate that the manipulation of financial performance can be beneficial if measures taken to address gaming result in other, more damaging, forms of manipulation.

This thesis focuses on performance measurement within the NHS and investigates whether changes in performance measurement during the study period, which involved reducing the scope for manipulation, led to beneficial or adverse consequences, with particular emphasis on managerial incentives and financial accountability. To investigate this central question two studies are undertaken. Study 1 investigates the incentive impact of a change in performance measurement system design on the achievement of the Government's strategic policy objective of delivering high service standards in the NHS cost effectively. A key feature of this change was an increase in measurement diversity. Study 2 focuses on the measurement of financial performance, which remained a key accountability measure throughout the period of study, most notably as a result of Trusts' statutory duty to breakeven. In relation to the central question, Study 2 investigates whether a form of officially recognised flexibility in reporting financial performance delivered benefits to patients and the public.

1.3 BACKGROUND TO THE THESIS

Over the last thirty years public sector organisations throughout the developed world, particularly Australasia, Europe and North America, have been subject to a wave of reforms aimed at addressing perceived issues of performance, efficiency and accountability. These reforms, captured by the notion of New Public Management, (Hood, 1991, 1995), have been characterised by quasi-market modes of governance, with their emphasis on contracting and competition between corporatised units of accountability, and by the adoption of private sector managerial best practice, including an increased emphasis on cost efficiency and performance measurement. The brunt of these reforms has been felt most in sectors such as healthcare where free markets are possible but where there has been a long tradition of Government intervention. In the English National Health Service (NHS) early moves towards market modes of governance, particularly in areas relating to accounting and accountability, were often characterised by the

retention of features grounded in previous hierarchical relationships. In this context, the continuing programme of reforms under successive political regimes, has fundamentally changed performance management and accountability relationships in the NHS and has given rise to an evolving framework of incentives faced by NHS managers. The aim of this thesis is, in the setting of English NHS hospital Trusts over the period from 2002-03 to 2007-08, to investigate the impact of the changing incentive framework, on the key foci of NPM reforms: performance, efficiency and accountability.

NHS Trusts were established in 1991 to own and manage public hospitals and to provide hospital services in a quasi-market relationship with healthcare commissioners and other healthcare providers. The objective was to create an 'internal market' where competition between Trusts would drive up the efficiency and performance of hospitals. The accountability of hospitals was also enhanced by making each Trust publicly accountable in its own right. This was achieved mainly through a requirement, established in statute, for each Trust to achieve financial breakeven. However, although NHS Trusts were quasi-independent self-governing organisations, elements of their previous hierarchical relationship with Health Authorities were retained. Rather than reporting direct to Parliament they continued to be performance managed by the Department of Health which in turn continued to be accountable to Parliament for the performance of the NHS as a whole, including NHS Trusts and the fulfilment of their statutory duties. Thus, whilst the introduction of NHS Trusts served to bring greater emphasis to agency relationships in an environment previously characterised by hierarchical modes of governance, it also introduced a dispersion and multi-layering of accountability within the wider NHS.

Fulfilment of a Trust's statutory duty to breakeven is determined by reference to the Trust's financial statements. However, during the 1990s, reported financial performance also acted as the primary measure by which the overall performance of a Trust was assessed and, in the absence of more reliable indicators, also became a proxy for Trust cost efficiency. Thus reported financial performance became the main reference point for measuring each of performance, efficiency and accountability. Reliance on this one measure was predicated on the effective operation of the 'internal market' which, in the event, failed to materialise. By the end of the 1990s, there was considerable concern that a reliance on financial breakeven, rather than leading to improvements in performance and efficiency, had

instead given rise to adverse consequences for patient care. The incoming Labour Government of 1997 identified that NHS service standards had become a major public concern (Department of Health, 2000) and proceeded to introduce a number of reforms which sought to align managerial interests with those of patients and the public by incentivising improved service standards as well as cost efficiency.

The key policy focus on service standards was realised through an evolving performance measurement regime which progressively reduced the scope for neglecting unmeasured activities in order to meet the key accountability objective of the statutory duty to breakeven. However, whereas the manipulation of performance measures which involved the neglect of unmeasured activities was viewed negatively, a contrasting manifestation of the emphasis on service standards was evident in the policy framework for the manipulation of financial performance. For most of the three decades leading up to 2007 some forms of manipulation of financial performance were viewed positively and were officially sanctioned where they allowed Trusts in financial difficulty time to address their underlying financial problems without recourse to strategies which might be damaging to patient care. The manipulation of performance measures and the different policy approaches adopted by the Department of Health to the manipulation of financial and nonfinancial performance provides an underlying theme for the investigations in this thesis.

The Labour Government's first step in the incentivisation of high service standards was taken when, in 2001-02, consistent with a wider international trend in public sector performance measurement, a multidimensional performance measurement system (PMS) was introduced into the English NHS, in the form of the Star ratings. However, the heavy weighting given to a narrow range of nonfinancial measures in the Star ratings, became a focus for disquiet about its vulnerability to manipulation. Rather than incentivising cost efficiency as a means of improving service standards, concerns continued to be expressed that unmeasured activities were being neglected and, further, that financial performance was being sacrificed in order to enhance Trust ratings. Given Trusts' statutory duty to breakeven, the potential for improving a Trust's Star rating by incurring a deficit in order to improve performance against nonfinancial measures was a particular source of concern for policy makers. In 2005-06 the Star ratings were replaced by the Annual Health Check, a system that was characterised by features aimed at further reducing the

scope for manipulation, most notably, increased measurement diversity and a more even balance between financial and nonfinancial measures.

Both the AHC and the Star ratings had a number of common features. Both were inspired by the balanced scorecard and, in both systems, aggregated performance ratings were awarded to Trusts based on their performance against a number of financial and nonfinancial measures. A limited number of these measures, reflecting key Government priorities, were both influential to the ultimate performance rating and common to both systems. Further, performance ratings were linked to a system of performance management which featured rewards for good performance and penalties for poor performance, including the dismissal of the CEO.

The Star ratings and the AHC, by reducing the scope for neglecting unmeasured activities, aimed to incentivise cost efficiency as a means of improving service standards. Further early emphasis was given to cost efficiency by the introduction, in 1998, of a new costing system which permitted the development of a measure of Trust cost efficiency, the Trust reference cost index. This index facilitated the benchmarking of Trust cost efficiency. In a further measure to strengthen incentives, to enhance accountability and to promote the sharing of best practice in service delivery, reference cost indices and Trust performance ratings were made publicly available.

Taken together, the NPM reforms of the early 21st century altered the incentive framework faced by NHS Trusts and reflected a move away from a focus on financial breakeven as the primary measure of performance, efficiency and accountability. However, by virtue of the statutory duty to breakeven, reported financial performance remained the primary measure for discharging the financial accountability of NHS Trusts. In the context of a more demanding performance measurement regime the achievement of this objective became increasingly challenging. For Trusts in financial difficulty, however, there was, until 2006-07, some flexibility in the reporting of financial performance from the system of ‘financial support’. In this system, surplus funds from elsewhere in the NHS could be transferred, on a non-recurrent basis, to Trusts in financial difficulty with the objective of allowing them to meet their immediate accountability objectives without recourse to strategies which would be damaging to patient services.

The manipulation of reported financial performance to achieve accountability objectives was deeply embedded in the institutional history and accountability of NHS hospitals. Financial support, with features reminiscent of budget reallocations, had its roots in the period when hospitals were managed as cost centres in a hierarchical relationship with health authorities. Although NHS Trusts were set up as quasi-independent, self-governing entities, elements of this hierarchical relationship were retained. They continued to be performance managed by Strategic Health Authorities (SHAs), the regional representatives of the Department of Health, who also oversaw and co-ordinated financial support transactions. Trusts in financial difficulty wishing to access financial support were required to agree a recovery plan with the SHA and this recovery plan was expected, in the majority of cases, to accommodate the repayment of the sums advanced.

The accounting policy for financial support treated both its receipt and repayment as adjustments to the revenue account. In this way, financial support, which acted to flatter both reported financial performance and Trust performance ratings, represented both a form of income shifting across the NHS, and a form of accelerated revenue recognition within Trusts receiving it. This accounting policy was in accordance with NHS accounting regulations which were in turn compliant with GAAP, adapted for the NHS setting.

In another reference back to hierarchical relationships, where hospitals were managed as cost centres, the operation of the system of financial support was largely hidden from public scrutiny. Prior to 2002-03, the financial statements of NHS Trusts were not required to disclose financial support transactions. From 2002-03 onwards however, following concerns expressed by the Audit Commission (Audit Commission, 2005), disclosures were mandated but were limited only to Trusts receiving financial support in the year of receipt. There were no equivalent requirements to disclose 'repayments' or for Trusts releasing surpluses for brokerage as financial support.

1.4 KEY FEATURES OF THE THESIS

The thesis focuses on the research setting of English NHS acute hospital Trusts and comprises two principal studies. Study 1 investigates the relationship between service standards and cost efficiency in the context of an evolving

performance measurement regime and Study 2 investigates the performance and accountability impact of financial support, an officially sanctioned form of flexibility in reported financial performance which, for Trusts in financial difficulty, permitted the meeting of financial objectives without sacrificing service standards.

A key feature of the research design for both Study 1 and Study 2 is the formulation of an objective function for NHS Trusts which draws, for its authority, on key Government policy documents such as the White paper, 'The New NHS' (Department of Health, 1997), the NHS Plan (Department of Health, 2000) and the Performance Assessment Framework (Department of Health, 1999). This objective function models performance against service standards as a function of cost efficiency in the presence of a financial resource constraint determined by Trusts' statutory duty to breakeven. In a setting which is characterised by an increased emphasis on agency relationships, this is then used, in both studies, as a basis of developing predictions which are testable using econometric analysis.

The use of broad based measures of nonfinancial performance and cost efficiency is also a key feature of this thesis. The measurement of both hospital efficiency and service quality is generally challenging. However, in the English NHS, the data collected for Trust performance ratings permits the construction of a more objective and broader measure of nonfinancial performance, uniform across Trusts and time, than has hitherto been possible, whilst the Trust reference cost index represents a broad based measure of the cost of a hospital's aggregate activity, adjusted for case mix and variations in cost due to location.

1.5 STUDY 1

1.5.1 Introduction

Study 1 in this thesis investigates the relationship between service standards and cost efficiency in the period following the introduction of a multi-dimensional, balance scorecard inspired, performance measurement system. It further exploits the juxtaposition of the Star ratings and the AHC to investigate whether the AHC, consistent with its additional features aimed at reducing manipulation, better incentivised cost efficiency as a means of improving service standards than the Star ratings.

In Study 1 the objective function provides a basis for investigating the impact of PMS design on the incentivisation of cost efficiency as a means of improving service standards. Two particular characteristics of PMS design are considered: measurement diversity and the balance between financial and nonfinancial measures. As an aid to the interpretation of this analysis a stylised 'performance comet' provides a diagrammatic representation of the impact of PMS design on the predicted distribution of nonfinancial and reported financial performance. Particular emphasis is given to the impact of PMS design on weak and inefficient Trusts which, from a policy perspective, is arguably of most interest. These Trusts comprise the tail of the performance comet and a consideration of this 'tail' is considered as part of the analysis. Production possibility frontier analysis, combined with the 'performance comet', then facilitates the generation of predictions for the impact of PMS design on the relationship between service standards and cost efficiency. These predictions provide the basis for the investigations in Study 1.

1.5.2 Contribution

The relationship between service standards and cost efficiency has been of considerable interest in the literature because of the concern that the pursuit of cost efficiency may be damaging to service standards (Mutter, Rosko, Greene and Wilson, 2011). However, studies of this relationship have been hampered by the lack of broad measures of hospital efficiency and of service quality. Prior studies have therefore defaulted to narrow measures of efficiency, such as theatre utilisation or patient length of stay, or measures of cost efficiency derived from the use of frontier analysis techniques (Worthington, 2004). These techniques have dominated the empirical archival literature on hospital efficiency but have been subject to considerable criticism because of the sensitivity of results to model specification and choice of input and output variables. On the other hand, narrow measures of efficiency such as theatre utilisation are limited in their ability to proxy for overall hospital efficiency.

Access to measures of hospital services has proved similarly challenging. Previous studies have therefore used patient and staff survey results or have used narrow measures, such as mortality rates which as a result of variations in case mix, are limited in their ability to measure quality. In the English NHS, however, the systematic collection and publication of data for the Trust performance ratings and

reference cost indices provides an opportunity for investigating the relationship between service standards and cost efficiency using measures which are arguably superior to those used in previous studies.

Both the Star ratings and the AHC, consistent with wider international trends in performance measurement, were inspired by the balanced scorecard. The Star ratings introduced, for the first time in the NHS, a systematic performance measurement system which incorporated nonfinancial measures as well as financial measures, with the objective of reducing incentives to sacrifice service standards in order to meet the statutory duty to breakeven. However, concerns persisted and in 2005-06 the Star ratings system was replaced by the Annual Health Check, characterised by greater measurement diversity and a better balance between financial and nonfinancial measures.

In the normative literature on the balanced scorecard Kaplan and Norton, (1992) argue that increased measurement diversity, featuring both financial and nonfinancial measures, leads to better performance. This argument has been supported by analytical research founded in principal-agent theory which has focused on the alignment of managerial incentives and which demonstrates that more (costless) measures are preferred if the additional measures provide incremental information about managerial activities of interest to the principal (Holmstrom, 1979; Banker and Datar,1989; and Feltham and Xie,1994).

A large number of empirical archival studies have investigated the performance impact of balanced scorecard systems in the private sector literature. However, studies adopting an empirical archival approach in the public sector accounting literature, particularly in the area of performance measurement and management, are notable by their absence (Goddard, 2010; Broadbent and Guthrie, 2008, 1992). In this literature, studies on balanced scorecard implementations have adopted a largely critical or interpretive approach and have focused on the tensions between different modes of governance, the political context in which performance measurement was introduced and the potential for manipulation.

The economics literature also contains a number of studies which investigate the impact of performance measurement systems in a healthcare setting. These studies adopt an empirical archival approach but rather than

investigating the relationship between service standards and cost efficiency they focus on the identification and consequences of trade-offs and the neglect of unmeasured activities. However, the results of two UK studies, which investigate the impact of Government targets on unmeasured aspects of care, deliver findings which are consistent with a positive association (Kelman and Friedman, 2009 and Propper, Sutton, Whitnall and Windjmeier, 2010). In these studies performance on unmeasured activities was found to be positively associated with performance on measured activities. The authors' interpretation of these findings was that the imposition of Government targets had incentivised process improvements which had led to greater efficiency and wider organisational benefits beyond the narrow objective of target achievement.

Study 1 makes the following contributions to the literature. First, it makes a contribution to the literature on hospital cost efficiency by conducting an investigation which uses broad based measures of Trust cost efficiency and service standards which are arguably superior to those used in previous investigations. Second, it contributes to the literature on the performance impact of PMS design by exploiting the quasi-experimental setting of the Star ratings and the AHC to investigate whether, consistent with Government policy objectives, the AHC was more effective at incentivising cost efficiency as a means of improving service standards than the Star ratings. Third, it contributes to the limited number of empirical archival studies on performance measurement in the public sector accounting literature in the UK and Europe, hitherto dominated by interpretive and critical studies.

1.6 STUDY 2

1.6.1 Introduction

Study 2 in this thesis investigates the performance and accountability impact of financial support, an officially sanctioned source of flexibility in reported financial performance which, for Trusts in financial difficulty, permitted the meeting of financial objectives without sacrificing service standards.

Study 2 applies the objective function of hospital Trusts to specify testable models and utilises financial support disclosures to conduct three investigations.

First, the impact of financial support on contemporaneous service standards is investigated by constructing a performance adjusted benchmark level of nonfinancial performance and comparing this with the actual level of nonfinancial performance for Trusts in receipt of financial support. Second, to test whether there are on-going benefits to service standards, consistent with improved business processes, the association between the receipt of financial support and the change in nonfinancial performance from the year of receipt to the following year, is also investigated. Finally, to investigate whether financial support allowed Trusts time to address their financial difficulties, the association between financial support and the subsequent change in underlying financial performance is also examined.

Descriptive statistics, in the context of a policy review are used to identify the impact of financial support on Trust and wider NHS accountability.

1.6.2 Prior literature and contribution

A large international literature exists on the manipulation of reported financial performance. This literature is dominated by the investigation of earnings management in the private sector in response to a number of incentives including those related to regulatory constraints, political costs and the avoidance of losses. Elsewhere a much smaller literature refers to the manipulation of financial performance using terms such as creative accounting and income smoothing. In this literature a number of studies investigate income shifting between group entities to deliver benefits to shareholders, mostly in the form of a lower tax burden, or to meet accountability requirements in institutional settings characterised by dispersed and multi-layered accountability. Vinnari and Nasi (2008) for example investigate income shifting between a Finnish utility company and its 'parent' municipal council which permitted the council to meet its breakeven requirement.

The empirical literature on earnings management views manipulation as having largely negative consequences. In the analytical literature however a more balanced view of the manipulation of financial performance is adopted and, under certain conditions, it has been shown to have positive consequences, for example, by allowing managers to devote productive effort in the interests of principals rather than adopting more damaging short term tactics in order to meet financial objectives (Demski, Frimor and Sappington, 2004; Arya, Glover and Sunder, 2003). However,

the literature also demonstrates that such flexibility should be made available only to 'good', highly skilled managers, who will use it to work in the principals' interests, and not to 'poor' managers who might squander the opportunity or use the flexibility opportunistically in their own interests. Demski (1998) further shows that accounting regulation can be used as a mechanism for selecting 'good' managers by making income manipulation difficult, requiring skill and hard work, especially to deal with the reversing out of its effects. Arya, Glover and Sunder (2003) also argue that it is the implicit role of regulators to make income smoothing challenging but not impossible.

In the UK, NHS Trusts have, since their inception in 1991, been required to prepare financial statements based on GAAP, as interpreted as being appropriate to the public sector setting by HM Treasury. The Department of Health then develops accounting policies which are compliant with the Treasury's requirements, including variations from GAAP to reflect the NHS context. The accounting policy for financial support represents one such adaptation which relaxes conventional conditions for revenue recognition thereby facilitating the achievement of immediate accountability objectives for Trusts in financial difficulty. The objective of this policy was to temporarily protect service standards by providing time for these Trusts to address underlying performance issues. In this setting the function of selecting skilled managers who will work in the public's interests is performed, not by accounting regulation, but by the Strategic Health Authorities (SHAs), the regional representatives of the Department of Health. SHAs are responsible for performance managing NHS Trusts, a role which encompassed, in a form of self-regulation, the oversight and co-ordination of financial support transactions. Trusts in financial difficulty wishing to access financial support were required to agree a credible recovery plan with the SHA and this often had to accommodate repayment of the sums advanced. Over the period of Study 2, however, concern was growing that, in the context of multi-layered accountability and an evolving incentive framework, financial support was being applied opportunistically to disguise deficits without underlying performance issues being addressed. These concerns were exacerbated by the lack of transparency regarding the impact of financial support on reported financial performance.

The adoption of private sector GAAP has represented a key NPM initiative in many international public sector reform programmes. Both GAAP itself and the

adaptations which have been deemed appropriate for the public sector context have been subject to critical review in the public sector accounting literature, which is dominated by a qualitative research paradigm. Adaptations have been subject to criticism on the basis that they erode understandability, comparability and accountability and that, in the context of a breakeven requirement, they can have implications for service delivery. Empirical archival studies however are notable by their absence from this literature.

Study 2 makes the following contributions to the literature. First, it contributes to the limited literature which investigates the manipulation of reported financial performance in a public sector setting in which accountability is dispersed and multi-layered. Second, it contributes to the literature on the manipulation of financial performance in the public sector by investigating the performance impact of the flexibility in reported financial performance it afforded to Trusts in financial difficulty. Third, it contributes an empirical archival study to the literature on the adoption and adaptation of GAAP by investigating the impact of a variation from conventional GAAP on both the performance and accountability of NHS Trusts.

1.7 FINDINGS

The findings from Study 1 indicate that the relationship between performance against service standards and cost efficiency is significantly positive throughout the period of study. This finding, which is based on the use of broad based measures of both cost efficiency and nonfinancial performance, confirms the findings of previous studies which use stochastic frontier analysis to obtain measures of hospital cost efficiency and which use, as measures of service quality, either survey based measures or narrow measures which are restricted in their scope, such as mortality rates. These findings provide evidence that the multi-dimensional performance measurement systems introduced by the Labour Government do provide a framework within which service standards are not, overall, sacrificed in the pursuit of cost efficiency. Further, the findings from Study 1 indicate that the relationship between nonfinancial performance and cost efficiency in the AHC is significantly stronger than in the Star ratings. These findings provide evidence that the additional measurement diversity and better balance between financial and nonfinancial measures which characterise the AHC reduce the

potential for manipulation and are consistent with the AHC being a more effective vehicle for incentivising cost efficiency as a means of improving service standards.

Study 2 investigates the performance and accountability impact of financial support. The findings from this study indicate that the policy of income manipulation represented by financial support delivered overall benefits to patients and the public in the form of better service standards in Trusts in financial difficulty. Further, consistent with the aim of allowing Trusts time to address their financial problems, there was also an improvement in subsequent underlying financial performance. However, there is also evidence that the system was not wholly effective during the study period. In a significant minority of Trusts, no improvements in service standards were observed. These findings contribute to the public sector literature on the manipulation of financial performance by demonstrating that manipulation can deliver benefits to patients and the public, but that the filters used for selecting 'good' managers need careful consideration, particularly in a setting characterised by multi-layered accountability and an evolving framework of incentives driven by a continuing programme of NPM reforms. A partial explanation for the findings from Study 2 may be the limited transparency surrounding the impact of financial support on key measures of public accountability. As a consequence public and parliamentary scrutiny of the policy, which could have served to mitigate the opportunistic use of financial support, was impeded.

The policy of financial support had a significant impact on the accountability of NHS hospital Trusts, even after the disclosure requirements from 2003. The amount of financial support in circulation in the period 2003-04 to 2005-06 amounted to about £1bn and was received by over a third of all Trusts. Further, of all Trusts which reported a surplus in the period of study, 35% had an underlying deficit and, on average, financial support transformed a mean Trust deficit of circa £8m to one of £2m. Despite the materiality of financial support transactions, disclosure requirements were imposed only on those Trusts receiving it in the year of receipt. There were no equivalent disclosure requirements for subsequent 'repayments' of financial support, nor of funds released for brokerage by Trusts in surplus. The system of financial support, as a form of income shifting, did not affect the overall reported financial performance of the NHS as a whole but it did have an impact on wider NHS accountability by reducing the number of NHS Trust deficits and breaches of the statutory duty to breakeven reported annually to Parliament by

the Department of Health. These findings contribute to the literature on the adoption and adaptation of GAAP in a public sector setting. The accounting policy for financial support, which represented an adaptation of GAAP considered appropriate to the NHS setting, largely concealed its impact on key measures of accountability and prevented effective scrutiny of the policy's effectiveness by Parliament and the public.

1.8 STRUCTURE OF THE THESIS

The remainder of this thesis is structured as follows.

CHAPTER 2 provides a historical analysis of the development of NHS hospitals, and their modes of governance, to provide a context for the development of the literature review in Chapter 3.

CHAPTER 3 commences in Section 3.2 with a review, within the wider public sector, of the development and characteristics of New Public Management, an approach to public sector management which draws on market based solutions and private sector best practice as a response to concerns about the performance and accountability of public sector organisations. It proceeds in Section 3.3 with a review of the literature on the cost efficiency of hospitals, a key concern of policy makers since the inception of the NHS. The literature on the role of multi-dimensional performance measurement systems, particularly those inspired by the balanced scorecard, in improving performance is then considered (Section 3.4), including studies concerned with the manipulation of performance measures. These initial sections of the Chapter focus on setting the context for the first study in this thesis, which investigates the relative effectiveness of two performance measurement systems in the incentivisation of cost efficiency as a means of improving service standards. Developing the theme of manipulation, the literature on the manipulation of reported financial performance is also reviewed (Section 3.5). This section of the Chapter as for other sections, considers both the private sector literature, which in this case is represented by the earnings management and income shifting literatures, and the public sector and healthcare literature. Lastly, the literature on the application of GAAP in public sector organisations is reviewed (Section 3.6). These last two sections of the literature review develop the setting for the second research question which investigates the extent to which the

manipulation of hospital reported financial performance can be beneficial to stakeholders such as patients, the public and Parliament.

CHAPTER 4 considers the institutional setting for the two studies of this thesis. Section 1, provides a brief description of NHS Trusts in the context of the programme of NPM reforms during the period of this study. Section 2 proceeds to consider the financial accountability of NHS Trusts and other NHS organisations; Section 3 covers the measurement of cost efficiency within the NHS. Then, in Section 4 a description of the two performance measurement systems, the Star ratings which was in operation from 2002-2005 and the Annual Health Check which replaced it in 2006, is provided. Finally, in Section 5, the policy of financial support is outlined with particular attention to its purpose, the accounting policy which was adopted and the consequent impact of Trust and wider NHS accountability

CHAPTER 5 covers the research methodology for both studies in this thesis. It commences with the generation of an objective function for NHS Trusts. This objective function provides a basis for the application of production possibility frontier analysis to generate predictions about the relationship between cost efficiency and service standards and on the relative impact of the Star ratings and the AHC on the incentivisation of cost efficiency as a means of improving nonfinancial performance. A stylised illustration of the impact of this analysis on the predicted distribution of the financial and nonfinancial performance of Trusts is provided in the form of the 'performance comet'. The research design determines the core data required and the Chapter continues with a description of the process of data collection and the construction of variables, particularly the construction of a uniform measure of nonfinancial performance across the study period. Finally, the Chapter concludes with the construction of non-parametric and multivariate models to facilitate the investigation of the two main research questions in this thesis.

CHAPTER 6 reports the findings of the two main empirical studies.

Finally, **CHAPTER 7** draws together the findings from the two investigations, interprets them in the wider context of NPM type innovations and discusses their implications for policy development. The contribution to the literature is reviewed, the limitations of the investigations considered and opportunities for further research explored.

CHAPTER 2 HISTORICAL DEVELOPMENT OF NHS HOSPITALS[2]

2.1 INTRODUCTION

This Chapter provides a brief introduction to the history of the development of hospitals and the role of accounting in that development, particularly in the context of the New Public Management initiatives introduced after 1979. Particular attention is given to issues of accountability, performance measurement and efficiency which represent the focus for the two studies in this thesis. These issues are explored in much greater depth in Chapter 4. However, this brief history and overview is provided with the objective of framing the literature review which follows in Chapter 3. This Chapter comprises the following: first, a brief history of hospitals prior to 1979; second, the development of New Public Management in the NHS in the period from 1979 -1997 when a Conservative Government was in power; and finally a review of the period from 1997-2008, after the election of a Labour Government.

2.2 HISTORICAL DEVELOPMENT OF HOSPITALS PRIOR TO 1979

Prior to the establishment of the NHS in 1948 there were approximately 3000 hospitals in existence, approximately 1000 operating in the more prestigious voluntary sector and 2000 operated by local councils. Voluntary hospitals, often founded by doctors, were completely independent and included internationally famous teaching hospitals such as St Bartholomew's, Guy's and St Thomas', smaller general hospitals and the least prestigious of this group, about 300 small specialist hospitals, specialising in particular types of patient, for example, children or particular types of treatment, for example, ophthalmology, orthopaedics, and neurology. Municipal hospitals in contrast were run by local authorities providing services originally covered by the provisions of the Poor Law and thus mainly catering for the poor, the elderly and the chronically ill.

[2] Numerous texts which deal with the history of the NHS have been used to source material for this short section. Much of the material is common across all texts and so individual points are not identified with specific texts. The texts used are: Klein, 2006; Mohan, 2002; Pollock, 2004; Rivett, 1998; Talbot-Smith and Pollock, 2006; Watkin, 1978 and Webster, 2002.

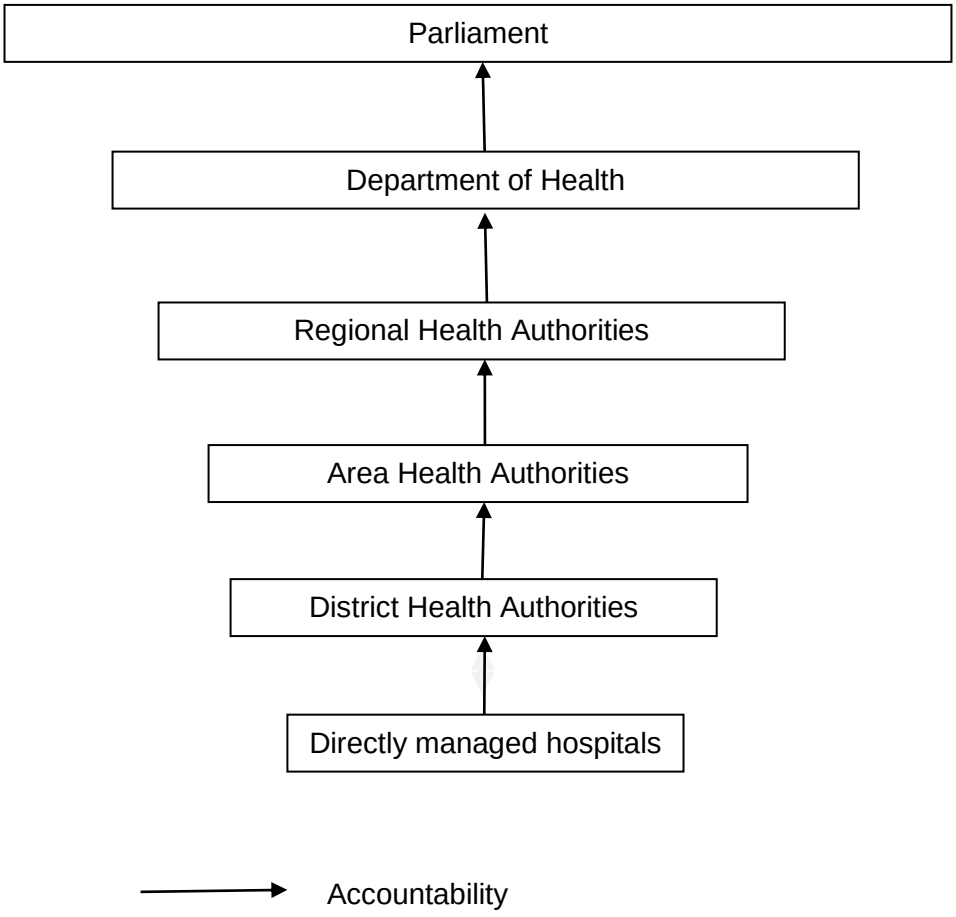
Distribution of hospital services thus developed in response to the strength of the voluntary sector and the priorities and financial resources of local authorities resulting in a patchwork of services which were variable in quality and quantity. In 1948 both types of hospital were brought together in a single system and an essentially collectivist model of provision was adopted with both financing and service delivery being provided by the state. Le Grand, (2003) refers to this period as being associated with 'knightly' assumptions about the motivations and behaviour of public servants and professionals. Those who operated the system were assumed to be selfless, altruistic and acting in the public interest. These 'knights' could be trusted to provide high quality services informed by their professional ethics and mores. Thus, under this 'trust' model (Le Grand, 2007, 2010) government provides the finance and sets the overall budget for a public service but no more. Those who provide the service are trusted to spend the budget in the public interest and to allocate resources so as to deliver a service that is efficient, responsive, accountable, equitable, and of high quality (Le Grand, 2010).

In contrast with these modes of governance the structure of the NHS during this period was hierarchical. Prior to the election of Margaret Thatcher as Prime Minister in 1979, there were 14 Regional Health Authorities, 90 Area Health Authorities and 192 District Health Authorities. Area Health Authorities covered areas which were broadly coterminous with local authority boundaries whereas District Health Authorities covered the catchment areas of district general hospitals. Funds flowed down to the District Health Authorities from the Department of Health, as illustrated in Figure 1, on the basis of a capitation formula. From this allocation hospitals negotiated a budget within which to operate. Management of the hospitals was performed directly by the District management team which was responsible for the performance of hospitals through functional lines of responsibility and professional regulation. Governance of hospitals during this period is thus most commonly associated with the notion of 'command and control' but as Le Grand (2002 a,b) observes, for much of this period there were 'few commands and precious little control'. Instead the delivery of services was dominated by the clan modes of governance of the medical profession (Lapsley, 1993; Bourn and Ezzamel, 1986; Ouchi 1980; Le Grand, 2002 a, b).

In this environment, governed by shared values and social norms, little value was placed upon the role of management and accounting information. Lapsley

(1993, 1992) goes as far as suggesting that, during this period, NHS financial managers operated by stealth, with inevitable consequences for transparency and accountability. Often, in the absence of sound management information systems, recourse to crude control techniques which focused on the management of real transactions, such as delaying payment to creditors, manipulating inventory levels and delaying the replacement of staff, was used to manage cash limited budgets.

Figure 1: Hospital services and the NHS in 1979



2.3 THE DEVELOPMENT OF NEW PUBLIC MANAGEMENT IN THE NHS: 1979 - 1997

By 1979 when Margaret Thatcher was elected, variable quality of care and waiting times for treatment led to an erosion of confidence in the 'trust' model which had also delivered standardised relatively low levels of service which patients were no longer prepared to passively, and gratefully receive. The belief in the 'knightly' motivations had become increasingly displaced by a view that 'knavish' motivations were more prevalent, and that, consistent with public choice theory, clinicians and managers acted in self-interest rather than in the public interest (Le Grand, 2003). Further, patients were no longer willing to be 'pawns' passively and gratefully receiving a paternalistic service that was arranged around the needs and convenience of the providers irrespective of their own needs and concerns. Following the election of Margaret Thatcher's government in 1979, driven by distrust of health service professionals and a belief that market mechanisms were the way in which user oriented, efficient services could be incentivised and delivered a programme of public sector reforms was instigated. These reforms, which looked to the private sector for managerial best practice, became part of the wider international sweep of reforms which became known as New Public Management (Hood 1991, 1995). In 1984, the first of a number of significant reforms was introduced. In response to criticisms of excessive delays in decision making (Lapsley, 1993, 1994) Area Health Authorities were abolished and hierarchies flattened. Further, a major challenge to the autonomy of the medical profession and their associated clan mode of governance came in the form of the Griffiths review (Griffiths, 1983) which introduced the concept of general management to the NHS. Griffiths, who was a businessman and therefore more familiar with hierarchical, rather than clan modes of governance, famously reported:

`In short if Florence Nightingale were carrying her lamp through the corridors of the NHS today she would almost certainly be searching for the people in charge.'

Following the Griffiths Report (Griffiths, 1983), the system of functional lines of responsibility and professional regulation with direct management of hospitals by the District Health Authority, was abolished and Unit General Managers, drawn from both public and private sectors, were appointed to manage hospitals and report through to equivalent positions in the Health Authorities. Hospitals remained as cost

centres within the primary unit of accountability, the Health Authority, and continued to be managed by reference to a budget allocation. Challenges to the ability of hospitals to live within their budget allocation continued to be dealt with through internal negotiation and reallocation of funds. LeGrand (2002a, b) and Lapsley (1993), interpret the culture of the NHS in the period following Griffiths, as one of transition from clan modes of governance towards an emphasis on hierarchical modes of governance associated with rules and bureaucracy, (Ouchi, 1980). Others, such as Pollitt (1986) and Pollitt et al. (1991) expressed concern that the new managerialism was too insensitive to deal with the political context within which public sector organisations operate and that the focus during this period was heavily weighted towards a perceived need to improve cost efficiency without sufficient regard being paid to other aspects of performance. The perception of these reforms, which were considered by many professionals in the NHS to have the potential to make a positive contribution to the performance of the NHS, suffered further damage when they became inextricably associated with the Government's aim of making cuts in public sector spending.

During this period the Conservative government also introduced other initiatives aimed at improving efficiency by economising on transaction costs (Coase, 1937; Williamson, 1975) and improving the use of resources. These included competitive tendering for non-core services such as laundry, cleaning and catering (Lapsley, 1993). These reforms, although relatively modest in their scope, were inspired by the notion of contestable markets (Bailey, 1981; Baumol, 1982; Baumol, Panzar and Willig, 1982) in which the market for public services is opened to potential new entrants as a means of driving up efficiency. This model found favour with the Government of the day as a more politically acceptable alternative to privatisation (Lapsley, 1993).

Much more significant reforms were to follow with the introduction in 1989 of the 'internal market', an idea promoted by Alain Enthoven, (Enthoven, 1985), an American adviser to the Prime Minister, Margaret Thatcher, during the 1980s. The NHS and Community Care Act, 1990 served to accelerate the move towards quasi-market modes of governance, characterised by contracting relationships, through the separation of the purchaser and provider functions. Purchasers, in the form of health authorities were allocated budgets to buy health care from 'providers'. Contracting between purchasers and providers was to be based upon the cost of

delivering services. However, hospital costing systems were insufficiently developed to allow such cost-based pricing to be established (Ellwood, 1995, 1996a, 1996b). In a reversion to hierarchical modes of governance contract negotiations defaulted to the negotiation of block contracts based on previous budget allocations. A further innovation aimed at enhancing market modes of governances was the establishment of 'GP[3] fundholders' who were allocated a share of the Health Authority budget allocation. GP fundholders were able to purchase services direct from providers rather than having the Health Authority purchase the services on their behalf. The original intention of the Government was that they would take over the role of local commissioning from the Health Authorities. However, the take up of fundholding status was patchy and, given the independent status of GP practices, the consequences for patients, particularly in relation to access to services, were questioned.

The separation of purchase from provision was institutionalised through the establishment of NHS Trusts as self-governing non-profit provider organisations, wholly financed by public capital, but delivering hospital services outside the formal boundaries of the Department of Health and Social Security. The non-profit status of NHS Trusts was an institutional feature which was embodied in a statutory duty to break-even (NHS and Community Care Act, 1990[4]). This statutory duty to break even represented the primary mechanism for the performance measurement and management of NHS Trusts and for the discharge of their public accountability throughout the 1990s. Other duties imposed by the Department of Health included a requirement to contain capital expenditure within an approved limit and not to exceed a cash flow limit. These additional requirements arose out of the alignment of the new NPM accruals based accounting of NHS Trusts with the traditional cash budgeting of health authorities where there was no distinction between capital and revenue expenditure.

In the context of block contracts and the volatility in expenditure arising from irregular payments, such as those associated with major capital projects, the

[3] GP: general practitioner

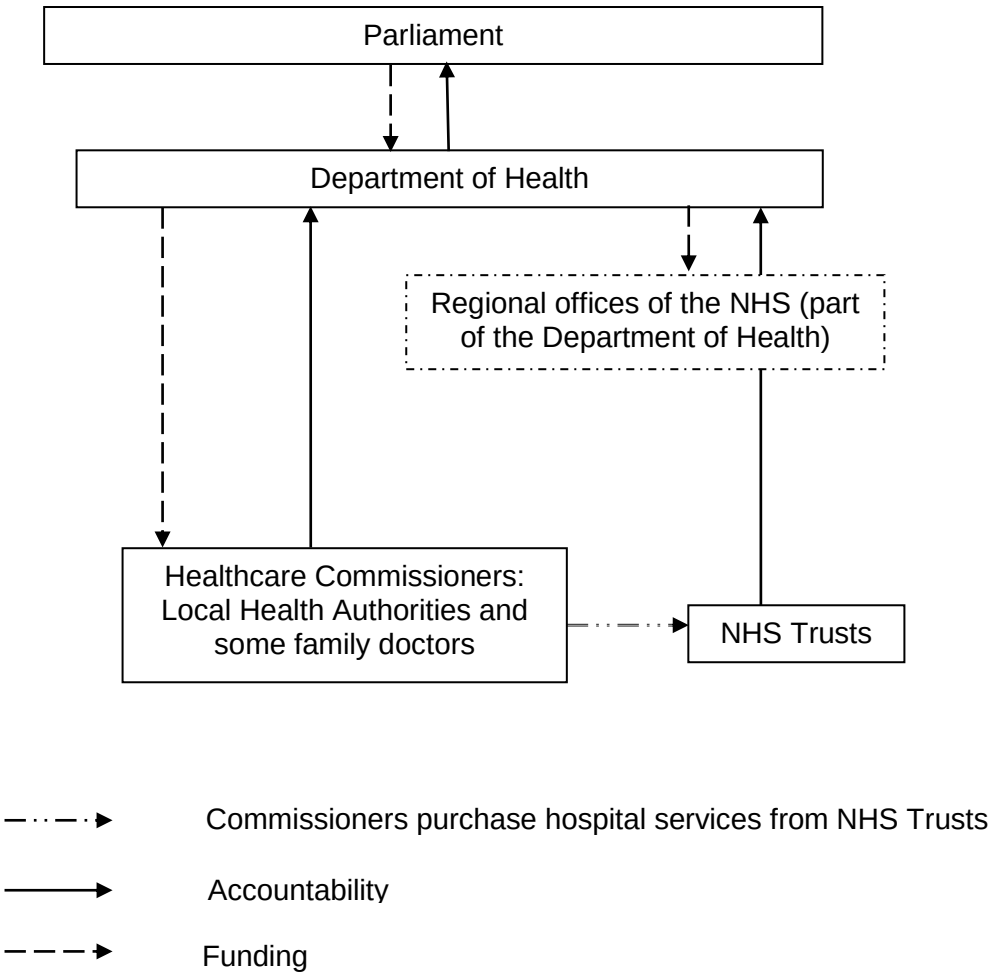
[4] National Health Service and Community Care Act (1990) available at: http://www.opsi.gov.uk/ACTS/acts1990/Ukpga_19900019_en_1.htm

achievement of the financial breakeven requirement proved challenging. In a further demonstration of inertia in the move from hierarchical to market modes of governance, a number of accounting mechanisms associated with the previous hierarchical regime were retained in order to give Trusts some accounting flexibility in the reporting of financial breakeven. The objective was to prevent the potentially adverse consequences for patient care that managers might otherwise adopt in order to achieve breakeven. These mechanisms included agreed transfers from Trusts' capital budget to their revenue account and a system of resource transfer, reminiscent of budget reallocations and a form of income shifting, which became known as financial support. The system of financial support, which is the focus for Study 2, is considered in depth in Chapter 4 which describes the institutional context for the two studies in this thesis. Trusts also used a variety of crude techniques based on the deferral of expenditure to a future period in order to assist in the achievement of breakeven (Lapsley, 1993), including the deferral of patient treatment. By the late 1990s however there was considerable public concern about waiting times for treatment and the impact this was having on patient care and service standards (Department of Health, 2000).

2.3.1 Summary of funding and accountability in 1997

Figure 2 illustrates the funding and accountability of hospitals throughout the 1990s until 1997, when the Labour Government was elected. Funds flowed from the Department of Health to local commissioners, the local health authorities and GP fund-holders, who purchased hospital services from self-governing NHS Trusts. NHS Trusts although operating outside the boundaries of the Department of Health, were accountable for their performance to the Department of Health through its regional offices. The Department of Health, in turn, was accountable to Parliament for all spending in the NHS which not included funds allocated to commissioners but also any overspending, in the form of reported deficits, in NHS Trusts.

Figure 2: Hospital services and the NHS in 1997



2.4 FURTHER DEVELOPMENT OF NEW PUBLIC MANAGEMENT IN THE NHS 1997-2008

Although the quasi-markets introduced by the Conservative Governments of the 1980s and 1990s had the objective of improving the efficiency and quality of services, making them more responsive to user needs and wants, the evidence regarding their success is mixed. There is some evidence that the efficiency of services increased, (Le Grand, 1999) but concerns remained about the quality and variability of treatment. Le Grand (1999) attributes the modest improvements and changes in the service to a failure to provide sufficiently strong incentives for improvement, and the imposition of constraints on the freedom of market

participants. For example, in London, where purchaser decisions threatened the viability of some London hospitals, (Klein, 2010), the Secretary of State intervened and placed restrictions on purchasing policies. Health Authorities were thus restricted from switching providers because such actions had the potential to destabilise existing providers and the political consequences of hospital closures were considered unacceptable. Similarly, patients had very little opportunity to exercise choice in either the purchaser of their services (the Health Authority or GP fund-holder) or the provider.

These issues were also identified by the new Labour Government, elected in 1997 which, in its review of the state of the NHS identified that that there was an absence of clear incentives to improve performance; that patients were disempowered and that services were not designed around the convenience and concerns of the patient. Further, a lack of national standards had led to unacceptable variations in treatment, (Department of Health, 2000, Chapter 2).

One of the first actions of the Labour Government was to abandon GP fundholding, which was associated with very high transaction costs. However the principle of the purchaser-provider split was retained. Local commissioning bodies were re-established in the form of Primary Care Trusts, with significant GP representation, and the regional offices of the Department of Health were replaced by Strategic Health Authorities. The main role of Strategic Health Authorities, as with the regional offices they replaced, was the performance management of local NHS bodies, including NHS Trusts which were accountable to the Department of Health for their performance, and the coordination of the strategic development of health services within their regional boundaries.

Following these structural changes, the Government went on to push through further NPM inspired reforms which simultaneously increased the prominence of quasi-market modes of governance whilst in other areas strengthened the culture of ‘command and control’ (Le Grand, 2002a). These policy initiatives were first outlined in the White Paper, ‘The New NHS’, (Department of Health, 1997), which made a commitment to improvements in both the quality and efficiency of NHS services. Indeed these were seen as being interdependent: higher quality services and improved cost efficiency went hand in hand (para. 3.8). In terms of quality, waiting times for treatment, in particular, had become the public’s top

priority (Department of Health, 2000), not only because of the length of time that patients were required to wait for treatment but also because of local and regional variations in waiting times. A more rigorous approach to the measurement of performance was to be developed, to be effected through a new system of efficiency measures (reference costs, para. 2.17) and through the establishment of a performance assessment framework (para. 3.14) with a wide range of measures covering fairer access to services, better quality of care and patient experience, as well as real efficiency gains (Department of Health, 1999). The effectiveness of the performance assessment framework was to be further reinforced by a linked system of performance management which incorporated, incentives for good performance, in the form of additional investment and freedom from regulation, and sanctions for poor performance, including direct intervention by the Department of Health. The CEOs of poorly performing organisation were also under threat of dismissal (Department of Health, 2000). The Performance Assessment Framework was trialled first in Health Authorities (Department of Health, 1999) and, a year later, the NHS Plan announced plans to develop and extend the framework to NHS Trusts (Department of Health, 2000). Moreover, in an initiative to improve Trusts' public accountability and to further incentivise good performance Trust performance ratings were to be made publicly available.

By 2002 the Government, having established its multi-dimensional performance measurement system (PMS) in the form of the Star ratings system, was signalling a return to further market-based solutions to promote improved performance and efficiency. These reforms were to be based on the notions of voice and choice, (Le Grand, 2010) and the contestability of markets, (Bailey, 1981; Baumol, Panzar and Willig, 1982) (Department of Health, 2002a, 2003b, 2005). Greater choice was to be offered to patients, including choice of both hospital and individual consultant, with cash following the patient to encourage utilisation of capacity. However, competition between providers was not to be on the basis of price but on the basis of quality. The price of each treatment was to be determined by reference to a national tariff (Department of Health 2002b, 2003a), a system known as Payment by Results. This new system, facilitated by the investment in the reference costing system which formed the basis for treatment pricing, facilitated the opening up of the market for hospital services to an increasingly diverse range of providers. Foundation Trusts, with greater commercial freedoms than NHS Trusts,

were to be established. These Trusts although non-profit and part of the NHS, were to be free of the statutory duty to breakeven, were able to borrow on commercial terms from banks and other lending institutions and had a greater degree of flexibility over their service portfolio. The intention was that over time, similar to the transition of hospitals to NHS Trust status, Foundation Trusts would become the dominant organisational form, with all hospital Trusts eventually becoming Foundation Trusts. During the period of this study, however, the majority of hospitals remained as NHS Trusts. Further, the market was to be made more open to private sector participation through the creation of independent sector treatment centres and through the commissioning of some services from private sector organisations.

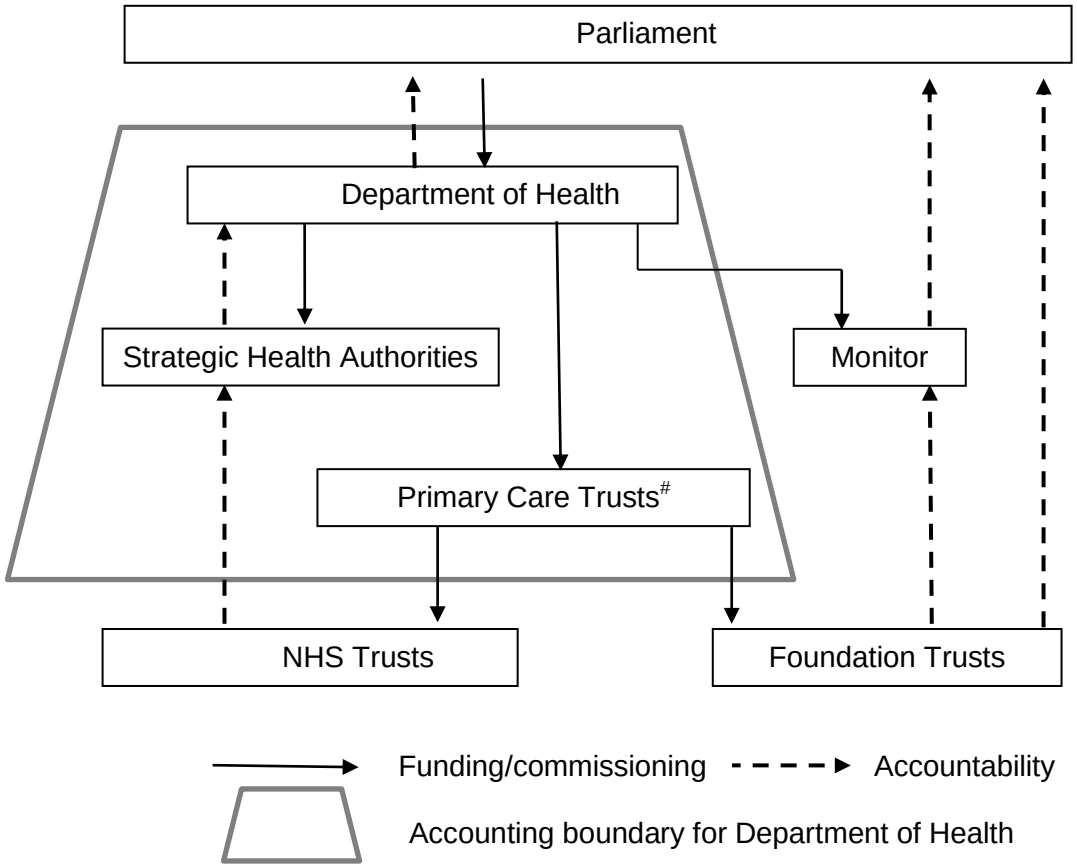
As a consequence of this series of reforms the accounting flexibilities that were carried over from the old hierarchical modes of governance were gradually whittled away. First, the flexibility accorded by the ability to defer patient treatment was eroded by the introduction of the Star ratings system with clear performance standards relating to waiting times. At around the same time the introduction of accruals accounting in all Government Departments, including the Department of Health, resulted in the withdrawal of the facility of transferring resources from the capital budget to the revenue account. During the period of this study therefore the system of financial support, in which surplus funds from elsewhere in the NHS were transferred into Trusts in financial difficulty as non-recurrent 'additional income' (NHS Accounting Manual[5]), became the most significant mechanism for according flexibility in the reporting of breakeven. The increasing momentum towards market-based governance in the period from 2002, finally resulted in the withdrawal of the system of financial support in 2006-07 following the introduction of Payment by Results and the opening up of hospital services to alternative forms of provider for whom favourable funding arrangements, as represented by 'financial support', were not available. (Audit Commission, 2006a, p.9).

[5] Available at: http://www.info.doh.gov.uk/doh/finman.nsf/ManualDownload?OpenView

2.4.1 The funding and accountability of NHS Trusts from 2001

The funding and accountability of NHS Trusts within the wider context of the NHS during the period of study is summarised in Figure 3.

Figure 3: Funding and accountability for English NHS hospitals



Adapted from: Audit Commission (2006a)

[#] A proportion of treatments were also commissioned from Independent Sector Treatment centres, which are regulated for quality of care by the Healthcare Commission (subsequently the Care Quality Commission).

NHS Trusts, Foundation Trusts and Monitor are all quasi-independent entities operating outside the boundaries of the Department of Health

Parliament annually votes funds to the Department of Health and these are largely devolved down to Primary Care Trusts on the basis of a capitation formula to fund the provision of healthcare for their locality. Primary Care Trusts commission hospital services from a variety of providers, mainly NHS Trusts and Foundation Trusts[6]. NHS Trusts, although self-governing bodies, are accountable for their performance to the Department of Health through the Department's regional intermediaries, Strategic Health Authorities (SHAs). An SHA's primary function is the performance management, on behalf of the Department of Health, of NHS Trusts and Primary Care Trusts operating within its area. A secondary function of SHAs is the coordination of the strategic development of health services in their region. A limited amount of funding flows to the SHAs to allow them to fulfil these performance management and strategic co-ordination roles. Both SHAs and PCTs fall within the accounting boundary of the Department of Health, whereas NHS Trusts and Foundation Trusts, as quasi-independent organisations contracting with PCTs to deliver services, fall outside it. This distinction is reflected in their financial statements which are prepared on different bases. Primary Care Trusts and Strategic Health Authorities, which are basically cost centres within the Department of Health, prepare an operating cost statement, the out-turn of which, for the purposes of public accountability, is compared with their 'Revenue Resource Limit', essentially their allocated budget. In contrast, both NHS Trusts and Foundation Trusts, which are regarded as 'trading organisations'[7] prepare an income statement, the balance on which provides the basis for assessing compliance with the statutory duty to breakeven. In contrast to NHS Trusts, Foundation Trusts are accountable directly to Parliament, each Foundation Trust being required to submit annual financial statements direct to Parliament, and are subject to regulation by an independent regulator Monitor rather than by Strategic Health Authorities. Monitor receives some funds from the Department of Health for the purposes of fulfilling this regulatory role. As a consequence of the different basis of regulation of Foundation Trusts, which permits greater independence and commercial freedom than NHS

[6] Primary Care Trusts, in addition to commissioning hospital services also commission primary care services such as general practice and dentistry services, which are largely delivered through private sector providers.

[7] See NHS Summarised Accounts from 2002-03 available at: http://www.nao.org.uk/publications/0304/nhs_england_summarised_accou.aspx

Trusts, they are excluded from the studies in this thesis. However, the establishment of Foundation Trusts from 2004-05 occurs during the study period and they have significance, particularly from a policy perspective, for the interpretation and discussion of the findings of the two studies.

2.4.2 Summary

In summary, during the period covered by this thesis, NHS acute hospital Trusts operated as quasi-independent organisations who contracted with health care commissioners (Primary Care Trusts) to deliver hospital services to patients. Trusts' statutory duty to breakeven represented the primary mechanism by which Trusts discharged their public accountability but from 2002-03 this was supplemented by a Trust performance rating system which aimed to incentivise cost efficiency as a means of delivering high service standards within the constraint of financial breakeven. The achievement of financial breakeven remained challenging for NHS Trusts throughout this period, particularly in the context of the new performance measurement system, which restricted the flexibility associated with the ability to defer patient treatment into a future accounting period. The only major flexibility available to Trusts during the period of study was the system of financial support whereby surplus funds from elsewhere in the NHS could be accessed by Trusts in financial difficulty. These funds served to flatter both reported financial performance and Trust performance ratings by being credited to revenue but were provided to Trusts who could demonstrate that they would use the funds effectively to support service standards whilst addressing their financial difficulties, often to the extent that the original funds advanced were repaid.

CHAPTER 3 LITERATURE REVIEW

3.1 INTRODUCTION

The establishment of NHS Trusts in 1991 represented a key element in the NPM innovations introduced to the NHS by successive Conservative Governments during the 1980s and 1990s. The increased emphasis on market-based governance mechanisms, particularly the separation of purchaser and provider and the emphasis on contracting, served to accentuate the principal-agent nature of relationships within the NHS (Hood 1991, 1995). Agency theory (Jensen and Meckling, 1976) predicts that agents will incur monitoring costs to reduce information asymmetry and to align incentives with principal's objectives. Consistent with this new emphasis, the 1997 Labour Government, in its role as executive agent to Parliament and the public, sought to align managerial interests in the NHS with those of patients and the public, by introducing a number of reforms aimed at incentivising cost efficiency and improved service standards. A new system for measuring Trust cost efficiency was developed and a multidimensional performance measurement system (PMS) introduced. To further incentivise Trust managers, a system of performance management, representing an extrinsic form of motivation also associated with agency relationships, was also introduced. This system was linked to Trust performance ratings with rewards for good performance and penalties, particularly the threat of dismissal, for poor performance. Further, the performance of each individual Trust was made publicly available, a measure which simultaneously served to align incentives and to enhance the public accountability of local hospitals within the wider context of the NHS.

These market based mechanisms were however introduced into an institution which had traditionally been associated with public administration and hierarchical models of governance. Indeed, many parts of the NHS, particularly those operating within the boundaries of the Department of Health, continued to be governed in this way. The alignment of market based modes of governance with the hierarchical modes of governance associated with public administration led to a complex multi-layering of incentives within the NHS and, in NHS Trusts, led to inertia in the relinquishing of mechanisms which served to deliver flexibility in the delivery of their public accountability, particularly the statutory duty to breakeven.

This Chapter reviews the literature relevant to these NPM developments and sets the context for the two studies in this thesis.

The Chapter commences in **Section 3.2** with a review of the development and characteristics of New Public Management. This review is set within the context of wider public sector reforms, going beyond the scope of the NHS alone, and includes a consideration of the literature which is critical of NPM initiatives. **Section 3.3** proceeds with a review of the literature on the cost efficiency of hospitals. This section commences with a brief review of the extensive, international, literature on hospital cost efficiency using frontier analysis techniques. It then proceeds to consider UK public sector accounting research into costing and cost efficiency before concluding with a short consideration of policy based papers drawn from the economics literature. Cost efficiency has been a key concern of policy makers since the inception of the NHS and improved cost efficiency has often been a key aim of NPM type reforms. Indeed, this aim of improving cost efficiency whilst improving service standards, was a key objective of the multi-dimensional performance measurement systems which were introduced into the NHS from 2002. The literature on the role of multi-dimensional performance measurement systems in improving performance, particularly those inspired by the balanced scorecard, is considered in **Section 3.4.** This section first considers analytical studies on multi-dimensional performance measurement before reviewing normative studies including those relating to the balanced scorecard. The impact of these models on performance is then considered, first from the perspective of empirical studies in the private sector and then from the perspective of the public sector, particularly healthcare.

A key focus of studies on performance measurement is that of manipulation. An extensive literature exists on this topic, particularly the manipulation of reported financial performance which has a substantial literature of its own. This is considered in **Section 3.5.** In the private sector this literature is dominated by the earnings management literature, but elsewhere there is a literature on income shifting between group entities aimed at realising overall economic gains in the form of a lower tax burden. Further, a small number of studies considers the manipulation of financial performance in response to incentives generated by institutional settings which vary from the traditional Anglo-American corporate model of governance and accountability. These include a Finnish study in which transactions between a

quasi-independent water company and a municipal City Council were manipulated in order to facilitate the fulfilment of the City Council's statutory requirement to breakeven. This section (3.5) on the manipulation of reported financial performance concludes with a critique of the methods used for detecting such manipulation and with a review of the literature which considers whether the consequences of manipulation are positive or negative.

NHS Trusts have a statutory duty to breakeven, a key component of their public accountability, compliance with which is determined by reference to their published financial statements. Their reported financial performance is determined by applying accounting policies that are compliant with UK GAAP insofar as it is relevant and applicable to the public sector. The application of private sector GAAP, itself an NPM type initiative, necessarily involves adaptations to fit the public sector setting. In the last substantive section of this Chapter (**Section 3.6**) the literature on the application of GAAP-based accruals accounting in public sector organisations, a considerable proportion of which is of a critical nature, is considered. The final section of the Chapter (**Section 3.7**) concludes with a summary of key issues arising from the literature review and identifies the contribution of this thesis to the different strands of literature considered.

3.2 NEW PUBLIC MANAGEMENT AND PUBLIC SECTOR ACCOUNTING RESEARCH

In the early 1980s concerns over the accountability and performance of public sector organisations led to a programme of reforms which became collectively known as New Public Management (Hood, 1991, 1995). In contrast with the public administration model which emphasised stewardship and which drew its management practices from the military (Hood,1991), New Public Management drew on market models and private sector best practice as mechanisms to deliver better cost efficiency and performance in public sector organisations. Despite being characterised by a degree of ambiguity, the value of the New Public Management (NPM) concept lies in its ability to capture key features and themes of the reforms that originated in the policies of the Thatcher era and continued throughout the period of the subsequent Labour government. These key features were summarised by Hood (1991, 1995) as: disaggregation of public sector entities into corporatised units; greater emphasis on cost efficiency; increased use of explicit standards and measures of performance; greater emphasis on output and results; the adoption of

management techniques which mimic private sector best practice; the increased use of contractual relations and an increase in competition for the provision of services; and finally, an increase in the application of responsibility accounting with clearer managerial assignment of responsibility and accountability for action. In a review of the sweep of reforms since the 1980s, Lapsley (2008) concluded that there was no foreseeable retreat from NPM reforms in the public sector and that, if anything, the momentum would continue.

The theoretical foundations of New Public Management lie in new institutional economics with its emphasis on efficient contracting informed by transaction cost economics, and rational, self-interested behaviour informed by agency theory (Groot and Budding, 2008; Hood 1991). New institutional economics, in its turn, drew from public choice theory to develop a number of administrative reform doctrines based on contestability, user choice and transparency (Hood, 1991). These, combined with an increased emphasis on 'professional' management and the adoption of business like management practices, gave rise to a wave of NPM reforms in public services in which market based mechanisms were adopted to address the perceived failures of hierarchical modes of governance. However these market based reforms have been criticised on the basis that the implications for public services have not been fully appreciated by policy makers. Whilst acknowledging that new institutional economics has relevance and value in the public sector, Ezzamel and Willmott (1993), for example, raise concerns about the erosion of professional values, clan based mechanisms of control and transactions underpinned by assumptions of trust (Ouchi, 1980). They argue that the increased emphasis on rational self-interested behaviour and a reliance on incentives to promote behaviour in the public interest, also paradoxically, gives rise to greater potential for dysfunctional responses and adverse consequences for the quality of public services. Indeed, the effectiveness of NPM type initiatives is considered to have been impeded by clinician resistance arising from the clash of cultures between the clinicians' traditional clan based modes of governance, the hierarchical managerial modes of governance associated with the Griffiths reforms (Griffiths,1983) and, latterly, market-based innovations (Lapsley, 2008). Pollitt et al (1986), Bourn and Ezzamel (1986), Preston, Cooper and Coombs (1992) and Jones and Dewing (1997) all find evidence of reluctance amongst clinicians to become engaged, for example, with the technique of budgeting with many clinicians

challenging the feasibility of reconciling the competing underpinning values and assumptions of clinical freedom and financial accountability.

Lack of confidence in New Public Management has also been eroded by instances of the adoption of private sector models which appear to have little substantive role to play in the management or governance of public sector bodies. For instance, Hodges, MacNiven and Mellett, (2004) find little engagement of the public with the notion of an Annual General Meeting for NHS Trusts, with the number of staff in attendance often exceeding the number of members of the public, despite considerable effort by Trusts to attract the public to them. They conclude that the NHS Trust AGM offers only a weak form of public accountability for NHS Trusts where lines of accountability are more strongly focused up through the Department of Health to Parliament. Other studies, which question the value of NPM initiatives that do not adequately recognise the public sector setting in which they are applied, include the questioning of the value of remuneration committees in the determination of NHS Trust Chief Executive pay (Ballantine, Forker and Greenwood, 2008b) and of the role of the Annual Report in the Australian fire service, (Kloot, 2009). This failure to adequately recognise the specific characteristics of the public sector setting has also been extensively criticised in the literature relating to the application of accruals accounting and generally accepted accounting practice in public sector entities as, for example, in studies undertaken by Barton (1999, 2000, 2004, 2005); Ellwood, (2003, 2008); and Ellwood and Newbury (2006)

Cultural resistance and poor application of NPM initiatives have both contributed to the perceived failure of some NPM initiatives but often unsuccessful models reappear in a different guise. Lapsley (2008) refers to this phenomenon as ‘Back to the Future’. He cites that the replacement of the, allegedly, failed Star ratings system by the Annual Health Check, which had many similar features, as an example of this ‘Back to the Future’ phenomenon. Other authors have gone further and suggested that, not only have some initiatives failed but the NPM movement as a whole has run its course, so much so that the movement is dead, (Dunleavy et al. 2005). New initiatives, it is argued, facilitated by an increasingly digital age, are being developed which run counter to NPM type initiatives based on disaggregation, competition and incentives. Examples of such initiatives include those which permit citizens direct electronic access and active management of their tax records, or

which allow patients to access their medical records and monitor their own treatment. These innovations represent a shift away from agency-centered processes, typical of New Public Management, to those which are citizen or stakeholder-centered, allowing citizens and businesses to substantially run their own interactions with government. However, de Vries (2010), in a review of Dunleavy et al.'s paper, argues that although NPM 'is in trouble … it is not really dead'. Further, Lapsley (2008) concludes that the NPM project, despite criticism of many its aspects, forges ahead with no immediate prospect of its demise.

Research into NPM initiatives has been extensive but despite its roots in new institutional economics, and the greater prominence of principal-agent relationships arising from, for example, the separation of the commissioning and provision of healthcare services, NPM research has been dominated by interpretive and critical studies using a qualitative or case study method (Goddard, 2010). This bias towards qualitative research is reflected in recent reviews of public sector accounting research such as Broadbent and Guthrie (2008) and Van Helden (2005). Broadbent and Guthrie's review builds on a previous 1992 study (Broadbent and Guthrie, 1992) by focusing on contextual or 'alternative' accounting research. It specifically excludes positivist studies and research published in US journals. Similarly in a review of public sector management accounting research Van Helden (2005) identifies that only 11 out of 55 studies (20%) used an economics based framework and only three out of 55 (5%) used an archival approach. As with Broadbent and Guthrie (2008, 1992) a consideration of US published research was excluded from his study. This bias towards qualitative research in the public sector accounting literature was further confirmed by Goddard (2010) in the first review of public sector accounting research which includes US studies. Goddard identifies the paradigmatic bias towards interpretive and critical research in Europe and Australasia and contrasts this with the US where research in Government and not for profit entities is dominated by positivist studies informed by neoclassical economics and utilising quantitative research methods which exploit readily available databases. A cross tabulation of author residency and methodological paradigm (excluding descriptive studies) showed that 95% of US papers adopted a functionalist paradigm as compared with 35% of papers from the rest of the world and a similar cross tabulation by research topic showed that 8% of US studies were on the topics of accountability and performance management, as compared with

41% for the rest of the world. These statistics highlight the bias towards qualitative research in the 'rest of the world' where the majority of research into public sector performance management and accountability has been undertaken. Goddard further raises the important point that, as a result of NPM initiatives, such as those related to performance measurement, a number of new and interesting databases have now become available in the UK but that these databases remain largely unexploited. He expresses concern that opportunities for undertaking investigations which could serve to improve our understanding of public sector reforms are being overlooked. Further, comparisons are drawn with the US where the dominant neoclassical paradigm in the top US journals and in the elite business schools has led to a marginalisation of behavioural accounting research (Williams, Gregory-Jenkins and Ingraham, 2006) which has been arguably detrimental to the overall field of study. The current paradigmatic focus in UK arguably poses similar potential risks to the development of public sector accounting research.

Notwithstanding the paradigmatic bias in Europe there have been a number of studies which have adopted an archival empirical method. Goddard identifies eight such papers out of a total of 188 (4%) in the period 2005-2007 and outside of this period examples exist in the area of NHS audit (Clatworthy, Mellet and Peel, 2002; Ballantine, Forker and Greenwood, 2008a) and in CEO pay (Ballantine, Forker and Greenwood, 2008b).

This thesis contributes to the limited research that adopts a positivist approach by utilising publicly available data to perform two empirical archival investigations into performance measurement and its manipulation in NHS acute hospital Trusts. The first investigation addresses a key policy concern: the relationship between service standards and cost efficiency and the impact of PMS design on the extent to which cost efficiency is incentivised as a means of delivering improved performance. The focus in this study is the relative impact of the Annual Health Check as compared with the Star ratings. The second study focuses on the manipulation of performance measurement systems, with particular emphasis on the manipulation of financial performance. Financial support, an NHS specific transaction, represented a form of income shifting between NHS organisations which largely served to accelerate revenue recognition in Trusts receiving it. For Trusts in financial difficulty it served to flatter both performance ratings and reported financial performance and to mislead Parliament and the public about the number

and size of NHS Trust deficits. The aim of this second study is to investigate the performance and accountability consequences of financial support and to determine whether it led to a public benefit in the form of better performance, both financial and nonfinancial, in Trusts receiving it.

3.3 COST EFFICIENCY IN HOSPITALS

3.3.1 Introduction

Cost efficiency has been a key policy concern for the NHS since its establishment in 1948 and a particular concern has been the efficiency of hospitals as they account for around two thirds of the NHS budget. The literature on hospital efficiency falls into three main categories. By far the largest strand is drawn from the economics literature and uses frontier analysis techniques to investigate and identify the determinants of hospital efficiency. This literature is international in its scope with studies set in over 30 countries. Second, a strand of literature exists in the UK public sector accounting domain. The focus of this literature has been on the development of costing systems, particularly the reference cost system which was introduced in 1998 and which provides the cost information for the Trust reference cost index, a measure of Trust efficiency. A third strand of literature is represented by studies which adopt a macro-economic perspective and which investigate the efficiency of the NHS overall. Each of these is considered in turn.

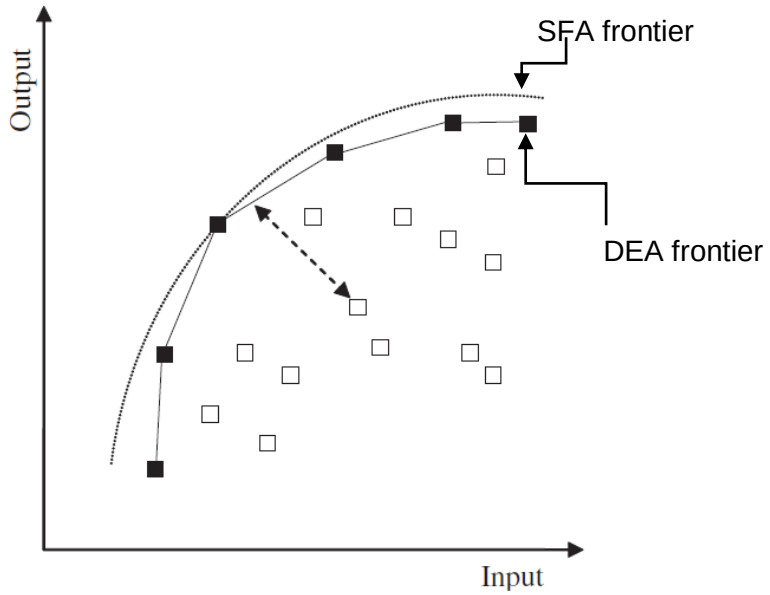
3.3.2 Hospital efficiency studies using frontier analysis techniques

By far the biggest stream of literature on hospital efficiency is represented by studies using frontier analysis techniques, such as data envelopment analysis (DEA) or stochastic frontier analysis (SFA), to obtain measures of hospital efficiency and to identify the factors which influence it. Over 700 such studies on healthcare organisations have been published in the last five years in over 30 countries, with the majority focusing on hospitals (Worthington, 2004, Hollingsworth, 2008).

Methodologically, frontier analysis techniques compare the efficiency of a hospital, not with a mean value as in econometric modelling, but with a frontier value. In DEA the frontier is determined by the best performing organisations and there are therefore organisations which appear as 100% efficient. In SFA the

frontier is a mathematically constructed frontier and organisations therefore rarely exhibit 100% efficiency. A stylised illustration of the two methods is shown in Figure 4.

Figure 4: Comparison of DEA and SFA approaches to frontiers



Adapted from Worthington (2004)

With DEA the frontier is constructed non-parametrically from the observations themselves, whereas with SFA a parametric function is fitted to the data. Organisations within the frontier (hollow points) are inefficient to the extent that they deviate from these frontiers. However, in both models the distance to the frontier (double dotted arrow) could be the result not only of inefficiency, but also because of misspecification of the production function or measurement error. There has thus been considerable concern about the sensitivity of results to model specification and the choice of input and output variables (Parkin and Hollingsworth, 1997; Chirikos 1998; Street 2003). Mixed findings from studies adopting different models have exacerbated these concerns. Further, the interpretation of the results from studies adopting frontier techniques has often proved problematic from a policy perspective. A number of factors, such as geographical location, specialisation and size for example have been identified as affecting hospital efficiency but the findings

so far provide insufficient information to be useful for determining policy. Street (2003), for example, points out that the apparent inefficiency of small hospitals could be for a number of reasons including the need for smaller hospitals to maintain a reserve capacity for dealing with accident and emergency cases which, proportionately, may be larger than for bigger hospitals. Addressing inefficiency based on these measures could therefore lead to adverse consequences for patient care. Resistance to frontier techniques has also been provoked by insensitive use of findings, particularly in the US, to deliver direct policy recommendations regarding budget controls and cuts in funding (Worthington, 2004). Many of the concerns regarding the use of frontier techniques centre around the fundamental issue that, in contrast to regression analysis, there has been no theoretical basis for identifying input and output variables, or of explaining the differences between one organisation and another (Ludwig, Van Merode and Groot, 2010). So, whilst the identification of the factors which influence organisational efficiency can provide interesting insights into an organisation's observed relative efficiency, many findings are difficult to interpret from a policy perspective.

Although the frontier analysis literature is international in its scope much of it is concentrated in the US. Three recent reviews of the literature illustrate this. The most recent (Rosko and Mutter, 2011) restricts itself to studies set in the US using the stochastic frontier analysis technique and published between 1994 and 2008, of which there were 27. The authors conclude that the application of stochastic frontier analysis to specific policy issues is in its infancy but that it has potential for development such that it could prove useful in appropriate contexts.

A slightly earlier review (O'Neill, Rauner, Heidenberger and Kraus, 2008) focused on DEA based studies over the period 1984-2004. Of 79 papers reviewed in a cross national study 49 (62%) were set in the US, whence the DEA technique originated. Only 6 were set in the UK of which two were set in Northern Ireland (McCallion, McKillop, Glass and Kerr, 1999; McCallion, Glass, Jackson, Kerr and McKillop, 2000) and one was set in Scotland (Parkin and Hollingsworth, 1997). All three studies which included England, (Jacobs, 2001; Maniadakis and Thanassoulis 2000; Maniadakis, Hollingsworth and Thanassoulis, 1999), covered the period prior to the Labour Government reforms of the early 21st Century.

Jacobs (2001) compared hospital rankings using the three cost indices used by the UK Department of Health in the period prior to 1997 with those obtained using DEA and SFA. She concluded that each frontier method has particular strengths and weaknesses, potentially measuring different aspects of efficiency. Further, she found that the differences in efficiency between Trusts were not large and concluded that the savings from bringing up poorer performers would be modest.

In two further studies, Maniadakis and Thanassoulis (2000) and Maniadakis, Hollingsworth and Thanassoulis (1999) used Malmquist productivity indices, computed using data envelopment analysis, to evaluate the efficiency of acute hospitals in the UK following the introduction of the internal market in 1991. They found that, even when service quality was incorporated into the analysis, productivity declined in the year following the introduction of the internal market but improved thereafter though not significantly.

The third most recent review of hospital efficiency using frontier analysis techniques was also an international study covering both DEA and SFA studies published from the mid-1980s until the early 2000s. Worthington (2004), rather than carrying out a comprehensive review, selected 38 studies using EconLit, the *Journal of Economic Literature* electronic database, to identify articles that were representative of frontier efficiency measurement in health care services. Of these, 20 (52%) were set in the US and only six in the UK, of which one was set in Scotland (Parkin and Hollingsworth, 1997); two concerned health authorities, the purchasers of healthcare services, (Thanassoulis, Boussofiane and Dyson, 1996; Giuffrida and Gravelle, 2001) and three investigated an acute hospital setting (Maniadakis and Thanassoulis, 2000; Street and Jacobs, 2002; Street, 2003).

Street and Jacobs (2002) and Street (2003) critically evaluate the OLS methodology used by the Department of Health for calculating the efficiency index, known as the casemix cost index (CCI), which preceded the Labour Government reforms of the early 2000s, with stochastic frontier techniques. Consistent with Jacobs (2001), they conclude that, using stochastic frontier techniques, hospitals exhibited a higher level of efficiency than was suggested by the CCI indices. However, Folland and Hofler (2001) and Street, (2003) warned that, although stochastic frontier analysis produced robust results for mean hospital efficiency, it

was an inappropriate technique for ranking individual hospitals and for setting targets because the individual estimates of hospital efficiency using SFA techniques are highly sensitive to study design.

One persistent concern of frontier based studies is the issue of quality. If quality cannot be observed and is a hidden component of costs then the pursuit of cost efficiency on the basis of frontier findings may lead to the perverse outcome of lower quality services. More recent studies using frontier techniques therefore adopt a two stage research design: measures of inefficiency are obtained through frontier techniques and then these measures are used as a variable in regression analysis. Deily and McKay (2006), for example, in a sample of urban acute care Florida hospitals, regress hospital mortality measures against a measure of cost efficiency, obtained from a stochastic cost frontier estimation, and find a positive association.

Further, in a response to the criticism that frontier based studies are not underpinned by theoretical analysis, Ludwig, Van Merode and Groot (2010), in an investigation of Dutch hospitals, use an agency perspective to predict and find a positive association between quality and efficiency. In this study a number of quality measures were obtained using staff surveys. A two stage analysis, where these survey based measures of quality were regressed against the outputs from a stochastic frontier estimation of cost efficiency, was then used to identify a positive relationship between quality and efficiency.

Ludwig, Van Merode and Groot (2010) used survey-based measures of quality rather than outcome measures, such as mortality rates, because, unless adjusted for case mix, outcome measures can be wrongly interpreted as indicative of low quality when high risk cases are being treated. The use of survey based measures of quality is also supported by Schwartz et al. (2011) who conclude that the value of survey based measures of quality has been underestimated. Their results suggested that such measures of quality were a more valid measure of performance than more objective measures and suggested that staff surveys could provide a useful complement to the use of patient surveys when evaluating organisational performance and the effects of improvement strategies.

In their recent study, Schwartz et al (2011), in a study of 470 US hospitals, also challenged reliance on individual metrics for quality, such as mortality rates, as

compared with composite measures. They identify that hospitals which are measured as highly efficient on a composite measure of quality, (as was determined for English NHS hospitals under both the Star ratings and the AHC), were rarely in the top quintile on most individual measures. They conclude that composite measures may be useful for policy purposes (for example, designing pay-for-performance systems) and for informing consumers about overall hospital quality, (a function which the Star ratings and the AHC performed), but they pose a challenge for hospital quality improvement programs. Although excellence in all components of the composite measure is desirable, hospitals may prioritise their improvement efforts to focus primarily on a few measures at the expense of others. The authors conclude that a balanced scorecard approach that employs both composite measures and individual component measures (as in the Star ratings and AHC) may be most useful for the development of effective, targeted strategies for incentivising the cost effective delivery of high service standards.

In the English NHS, although several studies exist which investigate the impact of the reforms of the 1990s, investigations into the impact of the Labour Government reforms since 2000 appear to be rather more scarce. A study published in 2006 for example was still focused on the market based reforms of the 1990s (Ferrari, 2006). However, a more recent study (Cooper, Gibbons, Jones and Macguire, 2010) investigated the impact on hospital efficiency of the competition reforms introduced by the Labour Government in 2006. Interestingly, this study did not use a frontier analysis approach to calculate hospital efficiency, but instead used average length of stay for hip replacement operations as a proxy measure. Regression analysis using a difference in difference estimator was then used to identify that hospitals affected by more competition reduced their patient length of stay more than those less affected. As the improvements occurred in the period prior to treatment they concluded that were genuine improvements in efficiency and not obtained at the expense of the quality of treatment.

In summary, a large literature exists which adopts frontier analysis techniques to obtain measures of hospital efficiency and to identify the factors that influence it. These techniques have so far not been adopted by policy makers because of concerns about the theoretical underpinning for the choice of input and output variables, because of the sensitivity of the results to model specification and choice of technique, and because frontier analysis is not yet sufficiently developed

for the purposes of measuring and ranking the efficiency of individual hospitals. The incorporation of controls for service quality into frontier analysis techniques has also proved a challenge and recent research studies have adopted a mixed method of frontier analysis and regression analysis to investigate these relationships.

3.3.3 Cost efficiency measurement in the NHS

A consistent policy focus since the establishment of the NHS in 1948 has been cost efficiency, particularly in hospitals which account for approximately two thirds of NHS expenditure. Many of the structural reforms of the NHS since its inception have been driven by the pursuit of cost efficiency, including the introduction of the 'internal market' by the Conservative Government in the 1990s. A key feature of the internal market was the separation of the purchasers and providers of healthcare which was effected through the establishment of NHS Trusts (Lapsley, 1994). The original vision was that the competition between hospitals would drive improvements in both efficiency and quality. However, as purchases of hospital services were to be made at cost, a key requirement for the operation of this market based system was the ability for hospitals to cost services effectively. Unfortunately little investment had been put into the development of costing systems (Ellwood 1995, 1996a, b) and they were found to be inadequate to the task. As a consequence the working of the 'internal market' was undermined (Lapsley, 1994) and contracting defaulted to block contracts, based on historical budget allocations. It was not until after the Labour Government was elected in 1997 that the necessary investment was made into a system, the reference costing system, that could be used as a basis for cost based pricing. This system was also used to develop a composite index of Trust cost efficiency, (Trust reference cost indices), based on the weighted average costs of individual treatments. Under this system a Trust with an average level of costs was given a Trust reference cost index of 100. Trust reference cost indices of more than 100 represented higher costs and lower cost efficiency and vice versa.

In the early years after its implementation the reference cost system was characterised by variations in coding and cost allocation practices which gave rise to considerable variation in reported treatment costs between hospitals. However, after substantial investment in the improvement of data quality, the Audit Commission (2010a) concluded that, of all sectors of the NHS, hospital Trusts had

the most sophisticated costing systems and the best quality data. However, for mental health and ambulance Trusts, where the roll out of the reference costing system occurred much later, data quality issues were more of a concern. Further, whilst acknowledging the need to continually work on improving the reference costing system, the Department of Health, in response to the data quality concerns reported by the Audit Commission also drew attention to the fact that these findings were based on survey and interview responses and that the perceptions of data quality appeared to be unduly harsh when compared with the results from a detailed analysis of the quality of reference costing in sixteen pilot sites (Audit Commission, 2010b). In reality data quality issues in acute services were concentrated in two specific areas: outpatients and non-admitted patient care (Department of Health, 2010).

Few studies have been carried out on the subject of reference cost indices and, in the main, they conclude that caution should be exercised in the use of reference costs for benchmarking. Two interview based studies in three NHS Trusts shortly after reference costs were introduced (Jones 2002) and again some years later (Guven–Uslu and Conrad, 2008), in which a range of clinicians and managers were interviewed, identified scepticism about the value of reference costs for benchmarking purposes largely because of concerns about data quality and the lack of alignment with measures of service quality. At the level of individual treatments (Health Resource Groups) there is no link between reference costs and measures of quality. As a consequence, benchmarking against cost alone could lead to an erosion of quality.

Similar issues were also identified by Northcott and Llewellyn (2003) in a study of Trust reference cost indices shortly after their introduction in 1998. Between 1999 and 2000 they observed considerable volatility in Trust reference cost rankings which they suggested was partly due to cost allocation and data quality issues. Further, they identified concerns over the use of ‘average’ as a benchmark, which is not a benchmark for excellence and which may or may not represent an acceptable standard of efficiency. Based on a study using critical discourse analysis (Llewellyn and Northcott, 2005), they concluded that as a consequence, hospitals were becoming more ‘average’ not only in their costs but also in their clinical activities. Thus, it was argued that the discourse of ‘excellence’ was being displaced by the discourse of ‘the average’. Further, Northcott and Llewellyn, (2005) raise

questions as to the ability of benchmarking exercises, such as the Trust reference cost index and performance ratings, to promote learning and improve effectiveness. Instead, they were viewed as being more likely to prompt a defensive response, designed to protect the status quo, for example, by asserting that the costs were inaccurate; that the mix of HRGs (Health Resource Groups) was skewed compared with national patterns, or that as a centre of excellence, more difficult and more expensive cases were treated (Jones,2002).

However, as predicted by Northcott and Llewellyn (2003) the Government remained committed to the development of reference costs and by 2009 the Audit Commission (2010), identified an increasing diversity of users and uses of reference cost data. Most significantly the Department of Health uses reference costs as a basis for the new hospital treatment cost based tariff system, Payment by Results (Department of Health, 2002b). Further uses were also identified. Both commissioners and providers used the data to support efficiency improvement initiatives and for service line reporting and management. The Department of Health also uses reference costs for calculating the weighted capitation formula which determines the level of funding passed annually to each PCT. Other users include academics and the National Office for Statistics which uses reference costs as an input to the macro-economic analysis of NHS efficiency (Department of Health, 2010). Further, the Department of Health, by annually publishing Trust reference cost data, has established itself as being at the leading edge of the development hospital cost governance metrics. As a consequence there has been international interest in the National Reference Costing Exercise (NRCE) from organisations such as the World Health Organisation, and the World Bank, and from countries as diverse as Albania and the USA (Llewellyn and Northcott, 2005).

3.3.4 Macro-economic studies on the efficiency of the NHS

Elsewhere, attempts have been made to measure the efficiency of public sector services, including the NHS, using macro-economic data. A recent study by the Office for National Statistics (ONS) (Phelps, 2009) reviews the productivity of public services over the period from the election of the Labour Government in 1997 up to 2007. Using a recently developed and experimental index of public sector output which included an adjustment for quality, (previously ignored), the ONS finds that public sector productivity fell by 3.4% over the 11 year period. This represented

an increase in the volume of input required to produce a unit of output of 0.3% per annum on average which compares with a figure of 0.4% per annum for health services. The higher loss of efficiency in health services was attributed to higher labour cost increases throughout much of the period which were offset, to some extent, by lower rises in costs elsewhere, as for example in drugs expenditure.

In an earlier study which focuses only on the NHS, Le Grand (2002a) used the Department of Health's measure of activity, the Cost Weighted Activity Index (CWAI), divided by a measure of expenditure adjusted for price changes, for hospital and community health services (HCHS) to obtain an index of the units of activity per unit of resources, a crude efficiency index. Using this measure Le Grand evaluates annual changes in NHS efficiency and concludes that efficiency rose until the Labour Government was elected in 1997 but declined in the three years to 2000.

The focus of both these studies is long term trends in productivity across the whole healthcare system. They do not attempt to identify the policy changes which may have contributed to such changes in efficiency.

3.3.5 Summary

In summary, the literature on hospital efficiency is dominated by studies adopting a frontier analysis methodology to obtain measures of cost efficiency and to identify the factors influencing it. However, despite continual development of this technique, considerable concern remains about the theoretical underpinning for the choice of input and output variables and the sensitivity of results to model specification. Further, a consideration of service quality in the context of studies using frontier analysis techniques has also proved a challenge. As a consequence, recent research has adopted a mixed method of frontier analysis and regression analysis, where the relationships between variables can be more effectively investigated. Two such studies which investigate the relationship between quality and cost efficiency, Deily and McKay, (2006) in the US and Ludwig, Van Merode and Groot, (2010) in the Netherlands, find a positive relationship. The investigation of the relationship between quality and efficiency using frontier techniques has further been hampered by the lack of availability of broad based measures of hospital quality. As a consequence studies have either resorted to narrow measures

of quality such as mortality rates (Deily and McKay, 2006) or broad measures based on survey responses (Schwartz et al., 2011; Ludwig, Van Merode and Groot, 2010).

Studies using frontier analysis to investigate the impact of policy reform on efficiency are limited and in the UK these are mostly focused on the internal market reforms of the 1990s. Only one study (Cooper, Gibbons, Jones and Macguire, 2010) investigates the impact of a post 2000 reform on hospital efficiency. However, this study, rather than adopting frontier analysis, uses regression analysis to investigate the impact of the Labour Government's competition reforms introduced in 2006 on length of stay for hip replacement patients. There are no studies which investigate the impact of the multi-dimensional performance measurement systems introduced by the Labour Government from 2002 onwards on NHS hospital efficiency.

In the accounting literature studies on hospital efficiency have focused on the development of costing systems and to the institutional responses to the introduction of Trust cost efficiency indices and cost benchmarking. These studies have adopted a mix of research paradigms with critical review and interpretive approaches being dominant. These studies are narrowly focused on the referencing costing system and so do not capture the interaction of incentives associated with a wider programme of NPM initiatives. There are no studies which consider the relationship between cost efficiency indices and other aspects of performance.

Finally a much smaller literature investigates trends in efficiency within the healthcare sector from a macro-economic perspective.

This thesis contributes to the literature on hospital cost efficiency in two ways: first, by investigating, using regression analysis, the relationship between cost efficiency and service standards in English NHS hospitals in the period from 2003-2008 and, second, by investigating the relative effectiveness of two performance measurement systems, the Star ratings and the Annual Health Check, in incentivising cost efficiency as a means of improving service standards.

3.4 MULTI-DIMENSIONAL PERFORMANCE MEASUREMENT

3.4.1 Introduction

The literature on multi-dimensional performance measurement is both extensive, multidisciplinary and covers both public and private sectors. This section, which reviews this literature, focuses on studies in the accounting and economics domains, and is organised as follows. First, formal analytical studies of multi-dimensional performance measurement (section 3.4.2) are considered. This is followed by a review of normative studies, particularly focusing on the balanced scorecard (section 3.4.3). Next, in section 3.4.4, the very large private sector literature which investigates the performance impact of nonfinancial measures and the balanced scorecard is reviewed. Section 3.4.5 proceeds by considering a number of critical studies which challenge the underlying assumptions of the balanced scorecard model and offer a partial explanation for the mixed findings from the empirical literature. Then, in section 3.4.6, a summary of this largely private sector based research is provided. Finally, public sector research into multi-dimensional performance measurement systems is reviewed and summarised in sections 3.4.7 and 3.4.8.

3.4.2 Formal analytical studies

Formal analytical studies of multi-dimensional performance measurement systems, which are drawn from the economics literature, adopt an agency perspective of multi-dimensional performance measurement and are sector neutral, being applicable to both public and private sector settings.

The limitations of financial measures for inducing managerial effort in the interests of principals are well documented. Financial measures are generally backward looking, do not capture aspects of performance of interest to principals, and can become divorced from an entities' strategic priorities. Proponents of the strategic performance management models, such as the balanced scorecard (Kaplan and Norton,1992, 1996a) argue that increased measurement diversity, in which both financial and nonfinancial measures feature, leads to better performance by reducing the potential for managers to neglect unmeasured activities and to trade one aspect of performance for another (Ittner, Larcker and Randall, 2003). These notions have been supported by analytical research founded in principal-agent

theory which has focused on the alignment of managerial incentives mainly in terms of reward systems. Holmstrom (1979), Banker and Datar (1989) and Feltham and Xie (1994) show that more (costless) measures are preferred if the additional measures provide incremental information about managerial activities of interest to the principal. Holmstrom (1979), however, observes that imperfect information on managerial actions is used extensively in practice to alleviate moral hazard and shows that any additional information, however imperfect, can be used to improve the welfare of both principal and agent. Concentrating on the role of management accountants in formulating performance measures Banker and Datar (1989) further identify a large class of situations under which linear aggregation, a simple and commonly used means of constructing performance measures, is optimal. They conclude that divisional performance evaluation on the basis of a profit measure alone will rarely be optimal. Feltham and Xie (1994) extend this analysis into a multitask environment and conclude that a diverse set of measures will be beneficial when an additional measure reduces either risk or non-congruity. These studies place their research largely in the context of reward systems which incorporate strong performance incentives. Although such contracts are rare in the public sector the introduction of NPM quasi-market modes of governance, with their emphasis on contracting and extrinsic managerial incentives, has led to an increasing accentuation of the agency aspects of relationships, notably in the UK. Further, despite the emphasis on the private sector in agency based accounting studies, the agency literature is not sector or entity specific in its application.

Jensen and Meckling (1976, p.5) define an agency relationship in very general terms as ‘a contract under which one or more persons (the principal(s)) engage another person (the agent) to perform some service on their behalf which involves delegating some decision making authority to the agent’. They point out that agency relationships, in which a utility maximising agent may not always act in the interest of the principal(s), exist in all types of organisation, not only private sector for profit entities but also in government organisations, mutuals and cooperatives and at all levels of management. The ‘problem’ of inducing an agent to behave in the principal’s interest, by the establishment of appropriate incentives and by the incurrence of monitoring costs, is thus a general one. A number of important contributions to the analytical and empirical investigation of performance measurement systems in public organisations using a principal-agent framework

have been made by Courty and Marschke who use data from a large scale job training programme in the US to investigate a number of research questions of interest in both the public sector and wider performance measurement literature (Courty and Marschke, 2003a, 2003b, 2004, 2007 and 2008). The use of performance measures was mandated in the federally funded agencies of the JTPA (Job Training Partnership Act)[8] and they were often published and linked to managerial payment (Propper and Wilson, 2003).

In a review of the principal agent multi-tasking literature, Courty and Marschke, (2003b) conclude that one of the main challenges in using performance measures is to manage gaming. They proceed to develop an evolutionary model of PMS design as gaming is revealed over time. This model suggests that because the gameability of a performance measure cannot be observed ex ante, the only way to identify a good performance measure is through experimentation. Courty and Marschke supplement their analysis by drawing on empirical evidence from the JTPA programme and observe that the performance measurement system had been updated several times over a period of over 20 years. This theme was further developed in a review of the evolution and development of performance measurement within the JTPA programme (Courty and Marschke, 2007) where they argue that in the public sector, which is characterised by objectives which are multidimensional and often not well defined, PMS designers cannot expect to anticipate all responses because local managers gain a superior understanding of how to influence the measures as they become accustomed to them. However, using the experience of the JTPA programme, they argue that the designers of government performance measures are capable of detecting how they are misaligned with programme objectives and taking remedial action. Gaming strategies need not therefore persist and, as system designers gain experience, performance measures may become more aligned with organisational objectives. This theme of continual review and constant renewal of performance measurement system design can be observed in many public sector systems. In the NHS, for instance, the Star ratings system was introduced to remedy the failures of past systems to address the trading off of service standards to achieve financial

[8] The Job Training Partnership Act (1982) was a US federal law enacted to establish federal training programmes for the young and unskilled to facilitate their entry into the labour market.

breakeven, a key measure of public accountability. This resulted, inter alia, in extended waiting times for treatment. The potential for manipulation of the Star ratings was in turn the subject of criticism (Bevan and Hood, 2006) and the system was replaced in 2006 by the Annual Health Check which incorporated features, such as greater measurement diversity, to mitigate the sources of manipulation that had been identified.

A further issue which Courty and Marschke investigated was the question as to whether the incidence of gaming responses aimed at maximising incentive awards affected organisational efficiency. They found (Courty and Marschke, 2004) first, that JTPA training agencies timed the reporting of trainees' performance outcomes in order to maximize incentive awards and second, through formal and empirical testing, found that these timing responses lowered the effectiveness of job training and gave rise to an adverse impact on organisational goals. They further demonstrate both formally and empirically that the existence of distortions in performance measures, which elicit gaming responses, can be identified by estimating how the association between a performance measure and the true goal of the organisation changes with the activation of the measure (Courty and Marschke, 2008).

As Courty, Heinrich and Marschke (2005) observe, the use of performance measurement systems and incentives in public sector programs has been a key component of 'reinventing Government' in the US and New Public Management initiatives in Europe and Australasia. In the UK, such initiatives have been driven by concerns about both performance and accountability and have resulted in accountability and governance structures which have been adapted from private sector best practice. This approach has served to accentuate the agency aspects of relationships amongst UK public sector organisations, with a much greater emphasis on performance monitoring and the development of stronger incentive frameworks. This developing institutional setting has thus opened up opportunities for the adoption of a positivist research paradigm and an agency based framework of analysis to investigate questions relating to the performance impact of multi-dimensional performance measurement systems in the public sector.

3.4.3 Normative studies and the balanced scorecard

The development of normative models for multi-dimensional performance measurement, particularly the balanced scorecard (BSC), which were initially designed for private sector application and later adapted to a public sector setting, are considered in this section.

Alongside the theoretical analysis of performance measurement a number of normative multi-dimensional performance measurement models, often aimed at the practitioner community, were also developed. These models, by incorporating a range of non-financial performance indicators in a normative framework, attempt to address the well-documented limitations of over-reliance on financial performance as a basis for performance evaluation, (Hopwood, 1972; Argyris, 1977; Otley, 2008). They include the Performance Prism, (Neely, Adams and Kennerley, 2002), the Performance Pyramid, (Lynch and Cross, 1991), the Business Excellence Model, (Kanji, 1998) and most notably the Balanced Scorecard, (Kaplan and Norton, 1992).

The Balanced Scorecard model aims to embed an organisation's choice of performance measures within the context of its strategic objectives and long term goals. The traditional balanced scorecard model comprises a number of measures in each of four key perspectives: customers, internal processes, learning and growth and financial performance (Kaplan and Norton, 1992, 1996a). The underlying premise is that the translation of strategy into specific objectives and measures linked in a causal chain of leading and lagging indicators in the four key perspectives leads to improved economic performance, measured either by accounting measures of performance, such as return on assets, or market based measures of performance, such as change in market value. Normative models such as the balanced scorecard draw on contingency theory with the choice of measures being influenced by the firm's strategy, competitive environment and its internal resources and competencies. The balanced scorecard thus says little about the number or identity of the measures, the weightings that should be applied or the timescale for monitoring purposes. Ittner and Larcker (2003) suggest that this silence may pose problems for practitioners attempting to apply the balanced scorecard in their own organisations and that this may offer a partial explanation for some of the perceived failures of the model. Other, more fundamental questions

about the balanced scorecard model have however been raised in a number of critical studies. These are considered in section 3.4.5, after a review of the literature which examines the empirical question of the performance impact of nonfinancial measures in section 3.4.4 below.

3.4.4 Empirical studies on the performance impact of nonfinancial measures in the private sector

A large accounting literature exists which investigates the performance impact of nonfinancial measures in the private sector. This literature is reviewed in this section and is further categorised into survey based studies using self-reported measures of performance, cross sectional studies using measures of actual economic performance and quasi experimental studies also using actual economic performance. Actual economic performance in these latter studies has been measured either in terms of accounting measures such as return on assets and, if a quoted company, in terms of market value.

The Balanced Scorecard was originally developed for the for profit private sector. Although increasingly adopted in not for profit and public sector settings, (see section 3.4.7.3), empirical archival research into the performance impact of the balanced scorecard, and more generally the use of nonfinancial measures, has taken place almost exclusively in the for profit sector.

In a review of empirical studies which test the association between the use of nonfinancial measures and economic performance Ittner (2008) identifies three strands of research. The first strand comprises large scale cross-sectional studies which investigate whether greater use of nonfinancial measures is associated with better economic performance. In this category two further subcategories are identified: studies using survey respondents self-assessment of performance and the second using actual measures of performance such as accounting or stock returns. The second strand of studies comprises quasi experimental company level studies examining whether performance improves after the introduction of nonfinancial measures. A further, more limited and recent, strand of literature tests a key assumption of the balanced scorecard model, that of cause and effect linkages between performance measures.

3.4.4.1 Survey based cross sectional studies using self-reported measures of performance

Ittner (2008) identifies that the majority of cross sectional studies support the hypothesis that the use of nonfinancial measures is associated with better performance. However the strength of the statistical relationship declines with the sophistication of the research design. Those studies which report the highest positive association are those which investigate self-reported performance rather than actual performance, as in Hall (2008) and Hoque and James (2000). This difference between perception and actual performance, also identified in Braam and Nijssen's (2004) investigation of the impact of the balanced scorecard in Dutch firms, raises questions about the quality of perceptual outcomes as indicators of actual economic outcomes. A partial explanation for the different results could be the well documented limitations of survey responses which are associated with common method bias (arising from the tendency of respondents to answer questions in a similar way) and the crude nature of categorical scales which leaves considerable scope for interpretation by respondents (Ittner, 2008).

3.4.4.2 Cross sectional studies using actual economic performance

Cross sectional studies using actual economic performance are fewer than those using perceptual outcomes. Recent studies include Ittner, Larcker and Randall, (2003) and Said, HassabElnaby and Wier (2003).

In an investigation of 140 financial services firms, Ittner Larcker and Randall (2003) investigated the association between measurement diversity and financial performance. The answers to survey questions were used to obtain three measures of measurement diversity: overall measurement diversity, financial measurement focus and non-financial measurement focus. Responses on ten performance categories were obtained: short term financial results (e.g., annual earnings, return on assets, cost reduction); customer relations (e.g., market share, customer satisfaction, customer retention); employee relations (e.g., employee satisfaction, turnover, workforce capabilities); operational performance (e.g., productivity, safety, cycle time); quality (e.g., defect rates, quality awards); alliances (e.g., joint marketing or product design, joint ventures); supplier relations (e.g., on-time delivery, input into product/service design); environmental performance (e.g., government citations, environmental compliance or certification); innovation (e.g.,

new product or service development success, development cycle time) and community (e.g., public image, community involvement). Each performance category was scored by survey respondents on a scale of 1 (low) to 6 (high) against a number of criteria: importance to long term success; extent to which goals were set for each performance category; measurement quality; and finally the extent to which measures in each category were used for (a) problem identification, (b) capital investment decisions, (c) performance evaluation and (d) external disclosure. Financial focus was measured by the values assigned to short-term financial results and non-financial focus by the average values assigned to the remaining nine performance categories. For the purposes of investigating the impact of measurement diversity on actual performance the scores for the extent to which the performance category was used for performance evaluation and decision making, had goals set and had high measurement quality, (i.e. all criteria except importance to long term success), were averaged. Overall measurement diversity was then determined as being the average score for all 10 performance categories; financial measurement focus was determined as being the average score for short-term financial results and non-financial measurement focus was determined as being the average score for measures related to the nine nonfinancial performance categories. To investigate the association between actual economic performance and measurement diversity four measures of actual economic performance were used: return on assets, sales growth, one year stock returns and three year stock returns. Using these measures, Ittner, Larcker and Randall found that overall measurement diversity and nonfinancial measurement focus had a positive association with one year stock returns. However there was no association between these measures and accounting returns or with three year stock returns. Similarly there was no association between financial measurement focus and any of accounting or stock return variables.

Said, HassabElnaby and Wier (2003) investigated the impact on performance when nonfinancial performance was incentivised through the inclusion of nonfinancial measures in performance contracts. Performance was measured as both return on assets and stock performance whereas a categorical variable of 0 or 1 was used to identify firms which used nonfinancial measures. A matched sample of firms was then constructed using industry, size and return on assets as matching criteria. They found that the use of nonfinancial measures was associated with

contemporaneous and future stock returns and with future accounting performance, but not with contemporaneous accounting performance.

Mixed findings have also been found in earlier studies which investigate the relationship between customer satisfaction measures and financial performance. In an investigation of 73 US retail bank branches Ittner and Larcker (1998b) found support for a positive association between financial accounting measures and a number of measures of customer satisfaction. Financial performance in this study was measured as business unit revenues, expenses, margin (revenue minus expenses) and return on sales (margin/sales). A positive association was observed for revenues, margin and return on sales but not for expenses

3.4.4.3 Quasi experimental studies

Quasi experimental studies, in contrast with cross sectional studies, trade generalisability for a more tractable research setting. In a study of eighteen hotels in a hotel chain, using monthly data over a period of seventy two months, Banker Potter and Srinivasan (2000) investigated whether, consistent with the normative balanced scorecard model, customer satisfaction measures were leading indicators of financial performance. Customer satisfaction was measured in two ways: complaints and the likelihood of return, based on customer responses on customer feedback forms. Financial performance was measured as revenue, costs and operating profit. Customer satisfaction was found to be positively associated with future revenue and profit but, consistent with Ittner and Larcker (1998b), not with costs. The lead time on the improvement was identified as being six months. However, as with Ittner, Larcker and Randall's (2003) choice of a single industry, the choice here of a single parent organisation resulted in a more homogeneous sample than multi-industry studies, mitigating the need, for example, to control for variations in incentive system, organisation structure, clientele and infrastructure. A limitation of this study was however the absence of a control group. The study relied instead on alternative control measures in the form of comparative data from competitor organisations, for example, an industry collected benchmark of average revenues per available room in the same location. The authors also acknowledged the essential arbitrariness and thus noisiness of the customer satisfaction scores. A second investigation in this study also found that the inclusion of nonfinancial

measures in managerial incentive contracts improved performance both on those measures and on revenues and profits.

In a second, more recent quasi experimental study, Davis and Albright (2004) compared the performance of four bank branches implementing a balanced scorecard with five control branches which did not implement a balanced scorecard. The measure of performance used in the study was a composite measure of financial performance based on nine key measures determined by bank strategists and BSC designers and which was linked to a bonus system. Comparing performance in one complete period prior to implementation and one complete period post implementation, they found that the performance of the implementers improved after implementation and that it improved more than non-implementers. The limitations of this study included the small sample size, the use only of non-parametric statistics and the limited generalisability of the results.

3.4.4.4 Summary of empirical studies on the performance impact of nonfinancial measures in the private sector

In summary, empirical private sector studies into the performance impact of nonfinancial measures have faced considerable research design challenges. In order to develop our understanding of these relationships a variety of innovative research designs characterised by increasing sophistication have been adopted. Not surprisingly the findings from such studies are mixed and although generally supporting a positive relationship between the use of nonfinancial measures and economic performance, the more sophisticated research designs deliver weaker results. Further, the results for the association between accounting performance and nonfinancial measures are generally weaker than for stock market performance. These results highlight that our understanding of these relationships needs further development.

3.4.4.5 Empirical studies examining the assumption of causal linkages between financial and nonfinancial measures

One of the underpinning assumptions of the original balanced scorecard model (Kaplan and Norton, 1992) was that improved economic[9] performance follows from the incorporation of nonfinancial measures into performance measurement systems. This 'cause and effect' relationship became further consolidated into the model as it developed from a performance measurement system into a strategic management tool in which performance measures are embedded in the firm's strategy through the modelling of cause and effect linkages amongst key performance indicators. This process became known as 'strategy mapping' (Kaplan and Norton, 2000). Failures in implementation of the balanced scorecard, by for example not giving sufficient attention to the task of causal mapping, were identified by Ittner and Larcker (2003) as giving rise to perceived failures in the BSC itself. However, although a large literature tests the association between financial and nonfinancial measures few studies go so far as to test whether a causal model was established or whether there were causal links between selected measures. In their study of financial services firms Ittner, Larcker and Randall (2003) use a survey instrument to investigate whether use of causal modelling as recommended by Kaplan and Norton (1996a) is positively associated with economic performance and found that it is not. Smith and Wright (2004), on the other hand, using path analysis to test an integrated causal model of performance in the US PC industry, find a positive association between customer loyalty and, through a chain of linkages, financial performance. First, customer loyalty was found to be positively associated with product price and sales growth, and in turn customer loyalty and sales growth were positively associated with return on assets. In contrast, Malina, Nørreklit and Selto (2008) using Granger causality analysis found only limited support for causal relationships between measures in a Fortune 500 company's balanced scorecard.

In summary, studies which investigate causal linkages between different aspects of performance have generated mixed findings. These findings are consistent with a developing strand of critical literature which challenges the

[9] Economic performance is defined, consistent with previous usage in this thesis, as being accounting performance, such as return on assets, or stock market performance.

underpinning assumptions and theoretical grounding of the balanced scorecard model. This literature is considered in the next section.

3.4.5 Critical studies of the balanced scorecard model.

Ittner and Larcker (2003) offer an explanation for the mixed findings of empirical studies which investigate the performance impact of the balanced scorecard by arguing that perceived failures in the model are the result, not of the failures in the model itself but of users failures to engage sufficiently with the disciplines it imposes, particularly that of strategy mapping. However, a number of studies have challenged this explanation by challenging the model itself and its theoretical underpinning.

In an early review of performance measurement research Ittner and Larcker (1998a) identify that our understanding of the notion of 'balance' in performance measurement, and particularly the balanced scorecard, is limited. They argue that there is not only a need to define more precisely the concept of 'balance' but also a need to understand better the ways in which 'balance' promotes performance. They note that a common view is that 'balance' is achieved through measurement diversity, as promoted for example by Kaplan and Norton (1992). Multiple measures in each of the domains of financial performance, customer related performance, internal business processes and learning and growth, it is argued, minimises the risk that valuable information relating to business performance is overlooked or lost. They further suggest that future research could provide additional insights into the concept of balance, by for example, shedding light on whether a wide selection of diverse measures or a smaller set of more reliable measures with greater predictive capability best promotes improved business outcomes. Ittner and Larcker (1998a) further note that the increased use of nonfinancial as well as financial measures exacerbates the problem of trade-offs, whereby actions taken to improve one measure may lead to short term declines in other performance measures. They argue that this creates a challenge for the retention of 'balance' in managerial actions and performance evaluation and conclude that additional research is needed on the inevitable trade-offs that managers make among financial and nonfinancial measures.

The issue of lack of precision, not only in the definition of 'balance', but also for example in isolating lead and lag relationships, is also raised by Nørreklit, Jacobsen and Mitchell (2008) in a critical review of the BSC. Further, in contrast to Ittner and Larcker, (2003), Nørreklit (2000, 2003) and Nørreklit, Jacobsen and Mitchell (2008), offer an explanation for mixed empirical findings, not by arguing that this is a consequence of users' failures to engage with the model, but challenging its theoretical basis. Nørreklit (2000) for example argues that the assumption of causality which underpins the model is flawed and not logical. She points out, by way of illustration, that customer loyalty does not necessarily lead to better financial performance if retained customers are those which are essentially unprofitable. In her subsequent paper, (Nørreklit, 2003), she goes on to argue that the theoretical underpinning of the BSC is weak and that its popularity owes much to the use of persuasive rhetorical devices characteristic of the consulting industry rather than the result of sound argument borne out of a strong theoretical framework. Nørreklit uses, as a basis for her analysis, Chapter 1 of Kaplan and Norton's (1996b) text, '*The Balanced Scorecard*'. Whilst acknowledging the academic credentials of the balanced scorecard, for which Kaplan won the 1997 American Accounting Association's award for best theoretical contribution, she also draws attention to the extensive use of analogies and metaphors and the use of emotive and loaded language. Nørreklit is particularly critical of the 'company-as-machine' analogy which Kaplan and Norton adopt when they liken a company to a jet aircraft and the CEO as its pilot. These analogies, she argues, oversimplify the task of managers, who are often operating in complex competitive environments, by reducing it to the need to press buttons and pull levers to achieve desired outcomes. This mechanical analogy, with references back to physical science, also underpins the cause and effect relationships which form the basis for the balanced scorecard, creating an illusion of scientific validity without providing solid evidence that such cause and effect relationships exist. Nørreklit also refers to the technique of contrasting the old (bad) with the new (good). So the old ways of measuring performance are bad, and the balanced scorecard which is new, is good. The balanced scorecard is associated with the new information age as compared with the old industrial age. 'Old' methods of performance measurement, using 'an accounting model developed centuries ago', contain the 'seeds of destruction' whereas the balanced scorecard is 'innovative and new' and recognises the value of intangible assets. Nørreklit concludes that, as a consequence of these rhetorical devices, the balanced

scorecard concept is open to interpretation, intuition and emotions, and is more akin to propaganda and the management guru genre than reasoned academic argument.

The absence of a theoretical underpinning for the balanced scorecard was also identified in Bessire and Baker's (2005) critical comparison of the BSC as compared with the much older French Tableau de Bord. Using a constructivist perspective they argued that this lack of theoretical foundation arises from an ambiguous and confused understanding of the political dimension of management control.

3.4.6 Summary of findings from the literature on performance measurement in the for profit sector

In summary the findings from empirical investigations in the for profit sector do offer support for the notion that nonfinancial measures improve economic performance, measured in terms of both accounting and stock performance. However, it is hard to avoid the conclusion that the findings do not paint a consistent picture. Possible reasons for these mixed findings include the absence of a sound theoretical underpinning; variations in research design; the absence of a cause and effect relationship between financial and nonfinancial measures (Nørreklit, 2000, 2003) and poor implementation and understanding of the balanced scorecard model (Ittner and Larcker, 2003).

3.4.7 Multi-dimensional performance measurement systems in the public sector

Again the literature in this section is drawn from both the accounting and economics domains. As identified by Goddard (2010) public sector accounting studies, including those whose research focus is the NHS, are dominated by critical commentary and interpretive studies. In contrast, economics based studies are generally empirical archival studies. The focus of these studies is however largely restricted to the identification of the gaming and manipulation of performance measurement systems. Both these streams of public sector research are reviewed.

3.4.7.1 Introduction

Notwithstanding our limited understanding of their performance impact, and despite being designed for application in the private sector, balanced scorecard inspired performance measurement systems have been increasingly applied in public sector and non-profit organisations. Kaplan (2001) essentially promoted a New Public Management type argument that non-profit organisations should mimic private sector management techniques in order to improve their performance. Whilst acknowledging that the financial perspective for non-profits acts as a constraint rather than a primary objective, Kaplan (2001), using case study examples to illustrate his arguments, proposed adaptations to the standard balanced scorecard model to fit the non-profit setting. Such adaptations included the positioning of the organisation's mission as the primary objective with elevation of the 'customer' perspective and the reduction in influence of the financial perspective. The need for financial viability and cost efficiency as factors enabling the achievement of non-profit objectives were identified as key factors informing the choice of measures in the financial perspective. A further adaptation was the interpretation of 'customer', which, in a non-profit setting was seen to have two dimensions: clients for the organisation's services and funders. He concludes by arguing that the balanced scorecard offers a strategy alignment tool that enables managers 'to align initiatives, departments and individuals to work in ways that reinforce each other so that dramatic performance improvements can be achieved.' Kaplan (2001) is essentially a normative contribution to the literature and no empirical evidence is provided to support the author's assertions.

The empirical literature investigating the impact of performance measurement in the public sector is more limited. A partial explanation for the absence of studies investigating the relationship between nonfinancial and financial measures, including cost efficiency, may be the challenges of research design in a setting where financial performance cannot be assumed to be the primary objective but, as Kaplan observed, is generally regarded as a constraint. Instead, studies whose research focus is the public sector concentrate on the dysfunctional consequences of performance measurement and are mainly to be found in the economics, rather than the accounting, literature. (Smith 1995; Propper and Wilson, 2003; Courty and Marschke, 2004; Bevan and Hood, 2006; Kelman and Friedman, 2009; Propper, Sutton, Whitnall and Windjmeier, 2010). These studies are largely

empirical archival investigations with the UK healthcare sector as the research focus. Exceptions are: Courty and Marschke (2004), in which the research focus is the US JTPA[10] programme, and Propper and Wilson (2003), which reviews the use and usefulness of performance measurement in both the UK and the US.

In the accounting literature, in contrast, the investigation of the application of performance measurement systems in the public sector is dominated by interpretive and case study analyses (Goddard, 2010). In their second review of 'alternative' public sector accounting research covering the period 1992-2006, Broadbent and Guthrie (2008) identified 38 papers out of a total sample of 452 on the topic of accountability, which included performance measurement and management, and governance. They observe that papers on performance measurement and management cover two main themes: the issue of whether organisations do what is expected of them and whether the systems imposed help to make transparent or obscure managerial action. Goddard's (2010) review, in contrast with Broadbent and Guthrie's (2008), included US journals and papers adopting a functionalist paradigm. He found that in the period from 2005 to 2007 the proportion of papers on performance management alone amounted to 10.6% of the sample of 188 papers, and accountability and governance accounted for a further 12.2%. However, despite the inclusion of US journals only 8 papers (4.3%), out of the total sample of 188, employed an empirical archival method as compared with 25% using interviews, 37% surveys and 10% case studies. Further, as has been previously identified in Section 3.2, on NPM and public sector accounting research, 95% of US papers adopted a functionalist paradigm as compared with 35% of papers from the rest of the world whilst, in contrast, only 8% of US studies were on the topics of accountability and performance management, as compared with 41% for the rest of the world. These statistics highlight the bias towards qualitative research in performance measurement and management.

The remainder of this section focuses on studies in the healthcare sector particularly the NHS and proceeds as follows. First, empirical studies on performance measurement in the NHS drawn from the economics literature are reviewed. Second, interpretive studies on performance measurement in the NHS

[10] Job Training Partnership Act

drawn from the public sector accounting literature are considered and finally, empirical studies in healthcare, including US based investigations, are reviewed.

3.4.7.2 Economics based studies on performance measurement in the NHS

In the economics literature investigations have focused on gaming responses to performance measurement. Evidence of gaming is found in both Bevan and Hood (2006) but both Kelman and Friedman (2009) and Propper, Sutton, Whitnall & Windjmeier, (2010), using more sophisticated statistical techniques, question whether the extent of gaming within the system leads to significant welfare losses.

Bevan and Hood (2006) in an interpretive review of performance measurement within the NHS argue that problems of both measurement and gaming were largely ignored in the Government's performance rating system for the NHS. By using limited data and some basic statistical analysis they provide prima facie evidence to support their argument that this was unjustified and that little reliance could be placed on improvements in reported performance. They cite, for example, references to ambulances waiting outside hospitals to reduce accident and emergency reported wait times and the use of trolleys as 'beds' to reduce emergency admission times. They go on to recommend improvements in the design of the system to reduce the incidence of gaming and conclude that the UK Government's performance rating system[11], consistent with Courty and Marschke's (2003) dynamic and evolutionary model of performance measurement system design, was an example of the Government 'learning by doing'. Overall, Bevan and Hood's review, although persuasive, was largely interpretive in its approach with only elementary statistical analysis to support the interpretation and conclusions. The persuasiveness of the paper however may be attributable in part to the use of rhetoric, including the generation of a sense of drama and excitement, and the use of metaphors and analogies, as identified in Nørreklit's (2003) review of the balanced scorecard. The authors for example liken the Star ratings to the Soviet system of 'targets and terror' and use words such as 'heroic' 'dramatic' and 'groan' for example as in:

[11] This system was known as the Star rating system, being based on a categorical assessment of performance ranked on a scale of zero to three stars.

a) '…governance by targets requires two sets of heroic assumptions..' (p.533)

b) '…there were indeed dramatic improvements in reported performance…' (p.533)

c) '…English health care system, in particular, groan(s) under regulation and audit'. (p.534)

In contrast, Kelman and Friedman (2009), in a statistical investigation of the impact of the accident and emergency waiting target in NHS acute hospital trusts found no evidence of gaming in the form of resource transfer (effort substitution) from unmeasured activities (in the form of resources from the orthopaedics department), and no evidence of a reduction in quality of care, (measured by death rates and re-admissions). Rather they find an improvement in care. They suggest two reasons for their findings: first, that complementarities exist such that process improvements to achieve access targets (waiting times for treatment) serve to improve overall hospital efficiency and so benefit other areas of clinical care and second, that negative feedback about gaming, from colleagues or from external stakeholders, could result in gaming being self-limiting. They go on to discuss possible managerial responses to limit the impact of dysfunctional responses: the inclusion of additional measures, adaptation of measures and, lastly, cultivation of the public service ethic amongst employees. Kelman and Friedman (2009) noted, for example, that resource transfer from orthopaedics to accident and emergency may not have been observed in their investigation because inpatient waits, for which orthopaedics was a key driver, were also subject to measurement. However, even if dysfunctional responses do occur, Kelman and Friedman argue that this does not of itself mean that the system fails to improve overall organisational performance. The appropriate comparison they argue is not between a PMS without dysfunctional responses and one with dysfunctional responses but between a PMS with dysfunctional responses and no PMS. Similarly, this could be extended to a comparison between two PMS, one with more dysfunctionality and one with less dysfunctionality. They conclude that, when presented with evidence of dysfunctional responses to performance measurement, abandonment of the PMS would not be an appropriate policy response.

This theme is further developed in Propper, Sutton, Whitnall & Windjmeier, (2010) who argue on the one hand that a weak incentive framework and multiple objectives allow managers to game performance targets but on the other hand, that

high profile targets, increased monitoring and publication of performance may be beneficial. High profile targets can act as missions around which employees coalesce and links to rewards can reduce the amount of effort devoted to tasks which employees value but which do little to improve social welfare. Further, as a result of inherent inefficiencies in the public sector, arising from multiple objectives and weak incentives, targets, particularly those which are focused on activities widely perceived as in need of improvement, can act to improve measured outcomes without diversion of activity or gaming. The authors then use difference in difference analysis to conduct a natural experiment which compares waiting times for treatment in English hospitals, where a performance measurement regime was in force, with those from Scottish hospitals, where there were no targets, to establish whether targets achieved their objective. They found that waiting times for treatment were significantly reduced, that there was no evidence of resource transfer from other activities and that quality of care was not eroded. Indeed, as in the case of Kelman and Friedman (2009) they found that the quality of care, measured by 30 day mortality rates, emergency readmissions and acute myocardial infarction, improved after the introduction of waiting targets. As with Kelman and Friedman they deduced that the process improvements necessitated by target achievement served to improve overall efficiency and had beneficial effects elsewhere. Some evidence of waiting list manipulation was detected but there was no evidence of welfare loss associated with this form of gaming. The authors did however acknowledge the limitations of their measures of quality of care and identified that these only related to patients who were actually treated.

The beneficial effects of performance measurement were also explored in another study which exploited the different regimes in England and Wales. Hauck and Street (2006) used four hospital Trusts serving both English and Welsh patients, three in England and one in Wales, to investigate the relative impact of the performance measurement regime in England on patient experience over the period of 1998-2004. Whilst acknowledging that factors other than the performance measurement regime may explain their results, they found that the English hospitals recorded increased levels of activity, undertook proportionately more day case treatment, and mortality rates fell. In Wales, in contrast, activity levels remained constant, the proportion of day case activity fell, proportionately more non-elective patients were admitted, and mortality rates rose. There was also partial evidence

that English patients faced lower waiting times than their Welsh counterparts and were more likely to be admitted within a target waiting period.

In summary, although prima facie evidence for the existence of manipulation and gaming exists in relation to the performance measurement systems applied to NHS hospitals, empirical archival studies provide evidence that such dysfunctional responses were not statistically significant across the whole population of hospital Trusts. Further, these empirical studies provide evidence that performance measures, based on waiting times, led to improvements in quality of care.

3.4.7.3 Performance measurement studies drawn from the public sector accounting literature

In contrast with the economics literature, the accounting literature has predominantly focused on interpretive and case study analyses of the implementation of multi-dimensional performance measurement systems, particularly those inspired by the balanced scorecard. By far the most dominant theoretical framework in the public sector accounting literature on performance measurement and management is institutional theory (Modell, 2009). Of twenty eight studies reviewed by Modell (2009), the research focus of three was healthcare and one of these (Chang, 2006) was set in the UK. This paper responded to the call from Brignall and Modell (2000) for more public sector performance measurement studies to adopt an institutional framework of analysis. Institutional theory, they argued, had been neglected within a large body of literature which focused on system design rather than system implementation and so typically ignored issues of institutional and political processes. They characterise managers' primary motivation for performance measurement as being legitimacy seeking rather than efficiency maximisation and argue that a framework grounded in institutional theory would facilitate a consideration of the dynamic interplay between funders, professionals and client groups. Developing this theme further Chang (2007, 2009) in two further studies concludes that performance measurement in health authorities, which commission rather than deliver health services, is used for legitimacy seeking purposes rather than as a strategic management tool (Chang, 2009) and that Kaplan and Norton's claim that the balanced scorecard model can be adapted for strategic management in the public sector fails to give sufficient

weight to the political context in which public sector organisations operate (Chang, 2007).

An alternative theoretical framework for interpreting the application of the balanced scorecard in a hospital setting was adopted by Aidemark (2001). In a case study investigation of Swedish hospitals Aidemark (2001) considers the meaning of a balanced scorecard (BSC) in a hospital setting by investigating the introduction of BSC systems in the hospitals of Jonkoping County Council. Using interviews and internal document analysis Aidemark examined the balanced scorecard implementation from the perspective of Ouchi's (1979 and 1980) markets, hierarchies and clan modes of governance. Scorecards were introduced in Jonkoping County Council in response to concerns over funding and the wish by clinicians to reassert patient needs within the performance measurement regime which had historically focused on financial performance to the detriment of other aspects of healthcare management. In this setting the conventional notions of balance, such as a balance between long and short term objectives, as expounded by Kaplan and Norton (1992, 1996a, 2001) were rejected in favour of a notion of balance as a network of perspectives operating in balance with no single perspective having priority. This contrasts with private sector BSCs in which there is an assumption that financial objectives are pre-eminent. In the non-profit hospital setting financial objectives were no longer pre-eminent and, consistent with Kaplan (2001) were interpreted in terms of constraints. Further, in this setting, in contrast with the original balanced scorecard model, the scorecard was not regarded as a top down tool of management control but as a mechanism for articulating a bottom up view of clinical activities and a means of constructing a new reality in which patient care was given priority. However, although the BSC allowed clinicians to reassert the patient care agenda in both management and wider political contexts, Aidemark concluded that the top down hierarchical characteristics of the balanced scorecard set in the context of the clan controls of professional clinicians created a mix of control mechanisms that was unstable and questioned whether the use of balanced scorecards in this setting would be a story of failure.

Other studies which consider the notion of balance include that of Ballantine, Brignall and Modell (1998) who note the influence of strategic objectives on the performance measurement systems of an NHS Trust on the one hand and a Swedish dental practice on the other. Given the need to tailor performance

measures to strategy, they consider the question as to what additional information is needed for a balanced performance evaluation and observe that in both the U.K. and Sweden there is a tension when attempting to balance the provision and use of financial and non-financial information. They conclude that the decentralised model of performance measurement in the Swedish dental practice was arguably more balanced than the centralised approach imposed on a UK NHS Trust. However, as in earlier private sector studies, (Kaplan and Norton, 1992, 1996a), the concept of 'balance' in this study remains imprecise.

A number of other studies, predominantly descriptive in nature, document and interpret the experience of healthcare organisations in adopting a balanced scorecard approach to performance measurement. Zelman, Pink and Matthias (2003) review the extent of balanced scorecard adoption within the US healthcare sector, including a notable example in the Duke Hospitals of North Carolina. Here the balanced scorecard was adopted at both an organisational level, as in the case of the Duke Children's Hospital (Meliones, 2000), and in individual departments, as in the Duke Women's Services Unit (Jones and Filip, 2000). An early study (Pink et al, 2001) reviewed one of the first system-wide adoptions of a balanced scorecard approach in 89 hospitals in Ontario, Canada. Inspired by a call for the use of balanced scorecards for performance measurement in healthcare organisations (Forgione, 1997) this 'bottom-up' application was driven by the need, in the context of new payment systems, to maintain service quality through efficiency improvements and increasing political pressure to improve accountability. A particular concern of those devising the system was the choice of financial measures and eventually four main categories were identified: financial viability, cost efficiency, liquidity and capital assets. The importance of data quality, the need for compromises driven by data availability, and the benefits that were perceived from the benchmarking of performance against other hospitals, were the key lessons learnt from this application. Similar results were found by Radnor and Lovell (2003a, b) who reviewed the application of the balanced scorecard in an English Primary Care Trust.

In contrast, a more interpretive paper (Agrizzi, 2008) draws on Broadbent and Laughlin's conceptualisation of performance measurement and management (Broadbent and Laughlin, 2009) to investigate a hospital's response to the introduction of multi-dimensional performance measurement in the English NHS.

She concludes that the hospital was deflected from its intended pathway of change and thus failed to meet intended outcomes despite having pursued both proactive and reactive strategies.

3.4.7.4 Empirical archival studies in the healthcare sector

Empirical archival studies in the healthcare sector tend to be concentrated in the US where Leslie Eldenburg at the University of Arizona has developed a stream of literature which exploits data from Californian hospitals. These hospitals are characterised by a variety of organisational forms in both the for-profit and non-profit sectors. The first strand of the literature investigates the impact of organisational form on performance (Eldenburg and Krishnan, 2003), on board composition (Eldenburg, Hermalin, Weisbach and Wosinska, 2004), and on the demand for accounting information (Eldenburg and Krishnan, 2008). The second strand investigates the response of hospital managers to changes in incentives arising from changes in accounting rules (Eldenburg and Vines, 2004) and medicare payments (Eldenburg and Kallapur, 1997; Kallapur and Eldenburg, 2005).

In the UK empirical archival studies in the healthcare sector are largely restricted to issues of governance (Ballantine, Forker and Greenwood, 2008a, b; Clatworthy, Mellet and Peel, 2002; Basioudis and Ellwood, 2005a, b) and financial reporting (Ballantine, Forker and Greenwood, 2007). An exception, concerning the relationship between financial and nonfinancial performance is an investigation of a small sample of NHS Trusts by Clatworthy and Mellett (1997) in which no relationship between return on capital and performance against patient charter standards was found.

3.4.8 Summary of performance measurement literature in healthcare

In summary there are few empirical archival studies in the healthcare sector which focus on performance measurement systems, whether in the US or elsewhere. Further, and perhaps as a consequence of the different institutional settings and the different paradigmatic traditions, the body of literature reflects, in many respects, the paradigmatic isolation identified in Goddard (2010): empirical studies are concentrated in US journals and settings and interpretive studies are concentrated in the UK and Europe.

Studies into the application of balanced scorecard performance measurement systems in healthcare organisations are largely interpretive and cover applications at both the individual hospital level and at a system wide level. The literature on system wide multi-dimensional performance systems questions, from a number of perspectives, whether such systems have the power to deliver desired outcomes because of the potential conflict between managerial and clinician modes of governance and because the wider institutional and political context in which such systems are applied do not appear to be adequately reflected in their design. These concerns have also given rise to questions about their ability to inform performance improvement strategies at the level of the individual hospital.

In the economics literature there is greater representation of empirical archival studies but these largely focus on the incidence of dysfunctional responses to performance measurement. However, the evidence from these studies, that there were improvements in the quality of care in English NHS hospitals which were subject to the new performance measurement regime, is consistent with a story of improved efficiency through process improvement which has beneficial effects throughout the hospital.

3.5 MANIPULATION OF REPORTED FINANCIAL PERFORMANCE

3.5.1 Introduction

Notwithstanding the development of performance measurement systems which incorporate a wide range of nonfinancial measures, financial measures remain of central importance to the evaluation of organisational performance, whether in the private or public sector. In the private sector financial measures represent the overriding focus for performance improvement. In the public sector, in contrast, although not regarded as representing the primary objective, financial measures remain important and as financial resources are regarded as a constrain, often include measures relating to the efficiency of the organisation (Pink et al, 2001; Kaplan, 2001). Measures of financial performance, however, are as vulnerable to manipulation and dysfunctional responses as other measures of performance. There is a long history of research exploring the manipulation of financial performance which dates back over 50 years (see, for example, Park, 1958). More recently this has evolved into a large body of literature known as the

'earnings management' literature. However, this term is very much associated with private sector studies. Elsewhere the literature also refers to the manipulation of financial performance using terms such as income smoothing and creative accounting (Amat and Gowthorpe, 2005). Further, there is a strand of literature which investigates the manipulation of subsidiary performance in response to economic incentives which deliver benefits to shareholders, such as those associated with the lowering of the overall corporate tax burden.

Underpinning much of the empirical research is an implicit assumption that the manipulation of financial performance has adverse consequences for stakeholders: earnings management seeks to mislead shareholders about the underlying performance of the firm and so results in inefficient resource allocation (Schipper, 1989; Healy and Wahlen, 1999; Dechow and Schrand, 2004). Regulation which seeks to restrict accounting choices and to improve the transparency of financial statements is therefore perceived as being beneficial. The opportunistic view of earnings management has, however, been challenged in the analytical literature which argues that a narrow view of earnings management may lead to policy initiatives which are detrimental to stakeholder interests (Arya, Glover and Sunder, 2003). This strand of the literature argues that earnings management contains additional information of value to principals, and that a certain degree of accounting flexibility is desirable when, for example, it incentivises goal congruent effort on the part of the agent (Demski, Frimor and Sappington, 2004; Arya, Glover and Sunder, 2003).

The remainder of this section proceeds as follows. First, empirical evidence for the manipulation of financial performance in response to incentives which are relevant to a public sector setting is reviewed (Section 3.5.2); secondly, the mechanisms used for manipulation and the methods for detecting them are considered (Section 3.5.3), and thirdly, the consequences of manipulation, positive and negative, are considered (Section 3.5.4). The section concludes with a summary section, section 3.5.5.

3.5.2 Incentives for the manipulation of reported financial performance

Much of the literature on the manipulation of reported financial performance has been conducted in a private sector setting. A significant proportion of this

literature identifies manipulation in response to incentives including those related to regulatory constraints, political costs, the avoidance of losses and executive compensation. The evidence provided by these studies is relevant to both public and private sector settings and is reviewed here. Being mostly set in the private sector the literature also identifies manipulation in response to incentives relating to the operation of capital markets. These include capital market transactions, such as share issues, which generally create incentives to raise the firm's share price; analysts' expectations, which create incentives to report steadily increasing earnings per share, and contractual outcomes, such as the determination of executive compensation and compliance with debt covenants, (Healy 1985; Healy and Wahlen 1999; Dechow and Schrand, 2004). This literature is excluded from this review.

However, incentives to manipulate financial performance do not operate only at the level of the ultimate reporting entity. A further strand of literature concerns the manipulation of the reported financial performance of subsidiary entities. The majority of this literature concerns incentives to shift income between subsidiaries in response to economic incentives in the form of a lower tax burden. However a small number of studies has identified manipulation of reported financial performance in response to incentives which are specific to an institutional setting which differs from the conventional Anglo-American, capital markets dominated, model of corporate reporting.

This section of the literature review proceeds to review the literature on the incentives to manipulate financial performance which are relevant to a public sector setting as follows: first, regulatory incentives; second, political incentives; third, loss avoidance; fourth, executive compensation. The last two parts of this section are informed by the view that, in the NHS, the system of financial support is a form of income shifting that allows Trusts in financial difficulty to meet their immediate accountability requirements whilst delivering benefits to patients in the form of better service standards. First, private sector studies investigating income shifting across group subsidiaries in response to incentives which deliver benefits to shareholders, mainly in the form of a reduced taxation burden, is considered. Finally, the manipulation of 'subsidiary' financial performance in response to accountability incentives which arise from specific institutional settings, which differ from the conventional Anglo-American capital markets model, is reviewed.

3.5.2.1 Regulatory incentives

Much of the research on regulatory incentives for earnings management is set in the financial services industry where failure to meet capital requirements incurs significant costs, associated with regulatory intervention. However evidence of earnings management to avoid these costs is mixed. Numerous studies find that, in the banking sector, loan loss provisions are used to manipulate earnings and/or capital maintenance ratios (as in Wahlen, 1994; Beatty, Chamberlain and Magliolu 1995; Bhatt, 1996; Lobo and Yang, 2001) but elsewhere no evidence is found (Ahmed, Takeda, Thomas, 1999; Collins, Shackleford and Wahlen, 1995). In a very recent study of over 11,000 banks over a period of sixteen years, Alali and Jaggi, (2011) find that, consistent with higher political and regulatory costs, large banks and those with high risk asset portfolios use loan loss provisions more than small banks and those with low risk asset portfolios to manage reported earnings. Further they found that, following the onset of the financial crisis, there was more earnings management in the years 2007-2008 than in previous periods. Although these studies are set exclusively in the banking sector such regulatory considerations often apply in the context of public services. In the English NHS, for example, Ballantine, Forker and Greenwood (2007) found evidence of earnings management in English hospital Trusts consistent with a statutory duty to report financial breakeven, failure to comply with which was reportable to Parliament, and had the potential for reputational damage and risk of dismissal (Ballantine, Forker and Greenwood, 2008b).

3.5.2.2 Political incentives

Private sector studies have also examined the influence of political costs on earnings management as for example when an industry is at risk of further regulation and high profits might bring unwanted political attention. Key (1997) for example found evidence of income decreasing earnings management in cable television companies during periods of US congressional hearings aimed at regulating cable prices. Similarly, Makar, Alam and Pearson, (1998) found that companies investigated for anti-trust merger violations engaged in earnings management to lower 'excess' profits to moderate evidence of an anti-competitive environment. Political incentives also feature in non-profit settings. Consistent with the predictions of Hoerger (1991), for example, Leone and Van Horn (2005) found that non-profit hospital CEOs manage earnings just above zero in order to achieve a

target level of earnings consistent with their budget constraint. Further, in the English NHS, the potential claw-back of surpluses was identified as an incentive for income decreasing earnings management in English NHS hospital Trusts (Ballantine, Forker and Greenwood, 2007).

3.5.2.3 Loss avoidance

In addition to earnings management to achieve analysts' forecasts there is a considerable literature which investigates the extent of earnings management used to avoid reported losses. This literature is predicated on the assumption that there are perceived costs associated with small losses which are avoided if a small profit is reported. However, although evidence of loss avoidance is provided in the form of discontinuities in the earnings distribution around zero, (as in Hayn, 1995 and Burgstahler and Dichev, 1997) the incentives for loss avoidance in the private sector, as compared with earnings decreases, are less well articulated. Burgstahler and Dichev, for example, support their hypothesis on the basis that 'References to the desirability of 'consistent profitability' are commonplace in annual reports, news releases, and press coverage, suggesting that there are incentives to avoid losses'. Further the use of earnings distributions to identify earnings management aimed at loss avoidance has been criticised on the grounds that there exist alternative explanations for the observed discontinuities around zero. (Beaver, McNichols and Nelson, 2007; Durtschi and Easton, 2005, 2009; Dechow, Richardson and Tuna, 2003). In the not for profit sector the incentives to avoid losses are more clearly articulated. Both Brickley and Van Horn (2002) in an investigation of US non-profit hospitals and Ballantine, Forker and Greenwood (2008b) in an investigation of English NHS hospital Trusts find that reported deficits are associated with relatively high levels of CEO turnover. Using research designs similar to those used in private sector studies they find evidence of the manipulation of reported financial performance to avoid reported deficits.

3.5.2.4 Executive compensation

Early evidence of the manipulation of accruals to increase earnings-based bonuses was found by Healy (1985). However the findings from more recent research, which uses more refined criteria, (Gaver, Gaver and Austin 1995; Holthausen, Larcker, and Sloan, 1995) are more mixed and are suggestive that other incentives sometimes act contrary to the incentive to maximise bonuses.

Gaver, Gaver and Austin's findings, for example, suggest that companies manage earnings to smooth earnings over a number of accounting periods rather than to maximise bonuses.

In the public and not for profit sectors, where multiple stakeholders and ambiguous objective functions, are associated with much weaker incentive frameworks the observed relationship between pay and performance, particularly financial performance, is weaker than in the private sector. For example, Brickley and Van Horn (2002) in an investigation of US hospitals found no difference in the relationship between pay and financial performance between for-profit and non-profit hospitals but observed that the relationship was weaker than elsewhere in the private sector. In contrast, in the UK public sector, Ballantine, Forker and Greenwood (2008b) found no relationship between CEO pay and performance in an investigation of English NHS hospital Trusts over the period 1998-2005. There is, however, considerable evidence that there exists a negative association between financial performance and CEO turnover, whether in the private or public sector (Brickley and Van Horn, 2002; Eldenburg, Hermalin, Weisbach and Wosinska, 2004; Ballantine, Forker and Greenwood, 2008b).

3.5.2.5 Income shifting between group entities to reduce the overall tax burden

Incentives for earnings manipulation can occur not only for group reporting entities but also at the level of subsidiary entities. A significant literature exists for example which investigates the shifting of income between subsidiary entities in response to economic incentives in the form of a lower overall taxation burden. Using disclosures on regional profitability from company annual reports a number of studies have found evidence consistent with income shifting from high to low taxation regimes (see for example: Collins, Kemsley and Lang, 1998; Oyelere and Emmanuel, 1998; Klassen, Lang and Wolfson, 1993).

A focus on specific types of transaction has also yielded evidence of income shifting. In a study of US multi-nationals, Newberry and Dhaliwal (2001), for example, find evidence that tax incentives influence where U.S. multinationals locate their interest deductions. The international bond offerings of US multinationals denominated in overseas currencies during the period 1987–1997, either through a foreign subsidiary in that country (yielding a deduction against foreign income), or through the U.S. parent or its U.S. financing subsidiary (yielding

a domestic interest deduction), were investigated. Consistent with income-shifting predictions related to the generation of economic benefits, it was found that bonds were more likely to be placed through a foreign subsidiary when there were binding foreign tax credit limitations that impaired the firm's ability to use domestic interest deductions, or when they are in domestic tax-loss carry-forward positions. Evidence was also found that bonds were more likely to be issued through a foreign subsidiary if the subsidiary was located in a country with generally high corporate tax rates than in the United States.

The transfer pricing of the exchange of goods and services between subsidiaries has also been identified as a potential mechanism for income shifting. In a study of transactions relating to research and development services, for which market prices are difficult to establish, Grubert (2003), using data from US Treasury corporate tax returns, found that transactions related to research and development intangibles represented a high proportion of income shifted from high to low tax countries. Further, the allocation of debt among subsidiaries and the shifting of research and development based intangible income together accounted for virtually all of the observed difference in profitability between high and low tax countries.

However, as is illustrated by Newberry and Dhaliwal (2001), the value of benefits derived from income shifting depends not only on differences in taxation regimes but also in the specific institutional features, such as the availability of relievable carried forward tax losses. Emmanuel (1999) also shows that the absence of a form of taxation, such as withholding tax, can have more impact on the value of benefits associated with income shifting than differences in headline corporation tax rates, which have formed the basis for many studies. Emmanuel (1999) also argues that the potential benefits associated with income shifting can also be mitigated by the tensions created by conflicting incentives specific to the institutional setting such as those associated with managerial compensation.

These studies illustrate that, in complex institutional and regulatory settings where incentives may be multi-layered, such as those associated with large decentralised and multi-national organisations, the importance of grounding research in a deep understanding of the institutional setting is critical to obtaining further insights into the incidence and impact of the manipulation of financial performance through income shifting.

3.5.2.6 Income shifting in response to other specific institutional incentives

Two further studies of relevance to the investigations in this thesis concern the manipulation of the performance of individual members of larger groups in response to specific incentives grounded in an institutional context which differs from the conventional capital market based Anglo-American model of corporate groups of companies. The first study is set in Japan (Thomas, Herrmann and Inoue, 2004). Until March 2000 the primary reporting statements for the Japanese stock exchange was not the consolidated financial statements but the parent company financial statements. After 2000, although consolidated statements became the primary reporting document for the stock exchange, parent company statements were also required to be submitted and they remained the only reporting statement required for the purposes of Japan's Commercial Code. For example, dividends and directors' bonuses are based solely on parent company reported earnings and both are prohibited where parent company earnings are negative, even though the consolidated financial statements show a profit. Financial analysts also provide forecasts of both parent and consolidated earnings. Finally, in Japan, there are strong cultural pressures to meet expectations and not be seen to fail. In this institutional setting, Thomas, Herrmann and Inoue, (2004) predicted that, in addition to consolidated earnings, parent company earnings would also be managed, through the manipulation of affiliate transactions, in order to meet or exceed a number of thresholds. These thresholds were those associated with loss avoidance, avoidance of a decline in earnings and avoidance of negative variations from forecast. Consistent with the manipulation of affiliate transactions, the manipulation of parent company earnings was found to be significantly higher than the manipulation of consolidated earnings. This study provides an example of an institutional setting, where managerial accountability is associated with parent company as well as group profits and which provides incentives not only for manipulation of consolidated earnings but also those of individual group entities.

The second study in this section is also set in a very specific institutional context in Finland. In a case study investigation of a Finnish City Council and its 'wholly owned' water utility company, Vinnari and Nasi (2008) describe how municipal accounting guidelines provided opportunities for flexibility in the transfer of funds out of the utility company and into the City Council which allowed the Council to fulfil its statutory duty to break-even. The City's Water Works was established in

1910 and was operated as part of the City's administration until 1994 when it was established as an 'independent' profit making enterprise wholly financed by City capital. The Finnish Water Services Act (119/2001) stipulated that water services charges should cover the running costs and investments of the enterprise over the long run and that the charges could include a reasonable return on the owner's capital investment (Compensation for Basic Capital). 'Reasonable' was not defined and varied widely across the Finnish water industry. Municipal accounting guidelines further determined that 'Compensation for Basic Capital' should not be treated as an appropriation of profit but as a financial expense. In the books of the City Council the amount received from the Water Utility was recorded as income and, as it was not ring-fenced for reinvestment in water services, was available to support general expenditure and to allow the Council to report financial breakeven. In the period under review the water company recorded profits very close to zero whilst paying out between 20% and 27% of revenue to City Council as 'Compensation for Basic Capital'. Further, the authors argue that, accounting treatment of 'Compensation of Basic Capital' as a financial expense served to disguise the extent of the transfer of revenues and undermined the accountability of both the Council and the water company. Councillors who 'like most users of accounting information…very obviously monitor and understand first and foremost the bottom line figure' (p.104) were misled about the underlying financial performance of the Utility company and of the Council. This case study illustrates how sector specific accounting policies can allow flexibility in the structuring of transactions to permit the satisfaction of regulatory requirements whilst simultaneously demonstrating how accountability can be impaired through the presentation of transactions even where there is full disclosure.

These studies offer novel insights into the manipulation of reported financial performance in response to incentives generated by an institutional setting which varies from the standard Anglo-American, capital markets oriented, model of financial reporting.

3.5.3 Mechanisms for the manipulation of reported financial performance

The manipulation of reported financial performance has traditionally been regarded as arising either from the discretionary manipulation of accruals, such as provisions for depreciation and doubtful debts, and from the manipulation of

discretionary expenditure, such as research and development, repairs and renewals and training, to achieve financial objectives. Whichever mechanism is used the manipulation of financial performance is conventionally regarded as having adverse consequences. Managerial accountability is impaired and resource allocation is based on misleading information. However, the manipulation of real transactions, such as decreasing or increasing research and development or marketing expenditure, is regarded as being particularly damaging as it has the potential for undermining the future strategic capacity of the entity. Dechow and Schrand (2004) go as far as to argue that '..such actions can have a significant impact on earnings quality and devastating effects on the company's future performance, and the transactions are a form of earnings management'. There are, however, few studies on the manipulation of real transactions because its detection can be problematic: intentional manipulation of transactions cannot easily be isolated from other, bona fide, reasons for variations in the level of expenditure. Nonetheless, Dechow and Sloan (1991), for example, showed that research and development intensive companies are more likely to cut research and development expenditure prior to the CEO's retirement. In the public sector, in the context of cash constrained budgets, the manipulation of real transactions has also been observed to be a frequently used tactic for meeting financial objectives. Lapsley (1993), for example, identifies that, when English hospitals were controlled via cash budgets, techniques such as deferring staff replacements were often used by NHS financial managers to prevent budget overspends. Interestingly, however, the US, and not the UK, provides the setting for a number of empirical archival studies which investigate the incidence of transaction manipulation in a hospital setting. Hoerger (1991) in an investigation of US hospitals over the period 1983-1988 found evidence that not for profit hospitals managed discretionary spending to meet their earnings objective. In a similar setting of US hospitals over the period 1990-2002, Leone and Van Horn (2005) also found evidence that non-profit hospitals adjust discretionary expenditure on charity care in order to manage earnings towards zero, consistent with their budget constraint. Eldenburg, Gunny, Hee and Soderstrom (2007) further found that, in Californian hospitals over the period 1997-2003, non-operating and non-revenue generating activities were managed to report positive income and that asset sales were managed by hospitals with incentives to decrease earnings to avoid regulatory scrutiny. However, using an alternative approach to the manipulation of real transactions, Kelman and Friedman (2009), in an investigation of English NHS

Trusts, found no evidence of effort substitution from the orthopaedics department to the accident and emergency department in response to an accident and emergency target wait time. Kelman and Friedman do acknowledge however that their measure for effort substitution, wait times in orthopaedics, was confounded by a waiting target for inpatient treatments for which orthopaedics wait times are a key driver. Similarly, however, Propper, Sutton, Whitnall and Windjmeier (2010) find hospitals met waiting time targets without diverting activity from other less well monitored aspects of health care and without decreasing patient health on exit from hospital. These studies all recognise the potential for adverse consequences for patient care from the management of real transactions to meet financial objectives. A research objective of the economics studies, in particular, was to identify whether such adverse consequences could be observed.

In contrast to the literature on the manipulation of real transactions, the literature on the manipulation of accruals is extensive, particularly in US capital markets based settings, where the literature is frequently referred to as the 'earnings management' literature. Accruals based earnings management is traditionally associated with the manipulation of year-end adjustments to accruals such as depreciation, doubtful debt provisions and inventory provisions. However, early studies of accruals based earnings management focused on the incidence and nature of accruals manipulation and found that revenue/receivables manipulation was commonly observed. For example, Dechow and Schrand (2004), in an analysis of 1745 SEC Accounting and Auditing Enforcement Releases identified 294 companies that had manipulated their accruals and identified the manipulation of revenue and receivables as being the most common form of manipulation, occurring in 70% of cases. This high incidence of revenue manipulation, which can involve the manipulation of real transactions with customers, is also supported by Nelson, Elliott and Tarpley (2003) who, in a survey of 253 auditors, found that over 50% of attempts at earnings management were identified as involving the adjustment of revenue.

More sophisticated statistical analysis of the incidence of accruals based earnings management is however hampered by the fact that it is, by its nature, hidden. Considerable effort has thus been put into the development of models to detect earnings management but, nonetheless, they remain subject to considerable criticism.

Jones (1991) developed an early model aimed at identifying abnormal accruals as the residual in a regression of total accruals against changes in revenue and the gross book value of property, plant and equipment.

$$ACC_{it} = \alpha_1 + \alpha_2.\Delta REV_{it} + \alpha_3.PPE_{it} + \varepsilon_{it}$$

Where:ACC_{it} = total accruals scaled by lagged total assets at time t for firm i; ΔREV_{it} = revenue in time t less revenue in time $t-1$ scaled by lagged total assets; PPE = gross book value of property plant and equipment scaled by lagged total assets; ε_{it} is an error term, identified as abnormal accruals.

However this model has been criticised for its failure to control for growth assuming, as it does, that all changes in revenue are non-discretionary. The inclusion of a modification to allow for changes in receivables (Dechow, Sloan and Sweeney, 1995) increases the precision with which the model detects the overstatement of revenue and, as a consequence, the discretionary component of accruals as represented by the residual ε_{it}. The 'modified Jones model' is as follows:

$$ACC_{it} = \alpha_1 + \alpha_2(\Delta REV_{it} - \Delta REC_{it}) + \alpha_3 PPE_{it} + \varepsilon_{it}$$

Where: ΔREC_{it} = change in receivables scaled by lagged total assets.

A further development in the aggregate accruals approach was to recognise that accruals reverse out over time, shifting or adjusting the recognition of cash flows. Using working capital accruals, which reverse out within one year, Dechow and Dichev (2002) measure accrual estimation errors as the residuals from a regression of changes in working capital on prior year, current year and one-year future cash flows from operations as follows:

$$\Delta WC_{it} = \alpha_1 + \alpha_2 CFO_{it-1} + \alpha_3 CFO_{it} + \alpha_4 CFO_{it+1} + \varepsilon_{it}$$

Where: ΔWC_{it} = change in working capital defined as change in accounts receivable plus change in inventory minus change in accounts payable minus change in taxes payable plus change in other assets, for firm i at time t ; CFO =

cash flow from operations; ε is the residual, a measure of abnormal accruals. In Dechow and Dichev (2002) all variables are scaled by average total assets.

Accruals and earnings quality is operationalised as the standard deviation of the residuals across a sample of firms, with a higher standard deviation indicating lower quality. As with the Jones' models this model does not isolate, in the residual, the components of discretionary accruals that are the consequence of intentional manipulation of earnings from those components which are due to bona fide estimation errors, for example, in the recoverability of debtors. Indeed, Dechow and Dichev (2002) argue that, even in the absence of intentional earnings management, accrual quality will vary with firm and industry characteristics with, for example, the volatility of operations affecting the accuracy with which accruals are routinely estimated. This failure to isolate intentional earnings management from the noise arising from routine estimation errors is common to all aggregate accruals models but, notwithstanding these limitations, aggregate accruals models are used extensively throughout the earnings management literature, including those studies set in the healthcare sector (Ballantine, Forker and Greenwood, 2007; Leone and Van Horn, 2005).

McNichols (2000) further argues that, it is difficult to estimate with confidence the extent to which accruals are associated with managerial discretion, whether intentional or otherwise, because of a limited understanding of the way that accruals behave. The Jones model for instance assumes a linear relationship between accruals and sales growth, yet there is no evidence to support this assumption. As a consequence, McNichols argues that the extensive reliance on these models in the literature is impeding further progress in the evaluation of earnings management behaviour.

An alternative widely used method of detecting earnings management is the examination of earnings distributions. This method has been most commonly used to investigate earnings management in response to a target, such as analysts' and managerial forecasts (Dechow, Richardson and Tuna, 2001 and Kasznik, 1999), and the avoidance of losses (Hayn, 1995; Burgstahler and Dichev, 1997 and Degeorge, Patel and Zeckhauser, 1999). Earnings management explanations for observed discontinuities in earnings distributions have however been subject to sustained criticism. (Durtschi and Easton, 2009, 2005; Dechow, Richardson and

Tuna, 2003; Beaver, McNichols and Nelson, 2007). Alternative explanations for the discontinuities include: sample bias, scaling issues and asymmetrical accounting treatment of tax gains and losses. In the latest contribution to this debate Durtschi and Easton (2009) show that market price per share, a commonly used deflator for earnings, does not have a linear relationship with earnings around zero giving rise to distortions in the observed frequency distribution. They conclude that the shapes of frequency distributions are not, of themselves, evidence of earnings management.

Another response to the limitations of aggregate accruals models is the development of a specific accruals method. This method requires researchers to identify specific accruals which are susceptible to manipulation, such as loan loss provisions in banks (Ahmed, Takeda, and Thomas 1999; Beatty, Chamberlain, and Magliolo 1995). However, the need to identify specific accruals demands considerable understanding of the institutional and regulatory framework within which entities operate and has resulted in a literature which is, at present, limited and focused on particular industries such as banking and insurance (Dechow and Schrand, 2004). This constrains the generalisability of the findings. However, notwithstanding this limitation, McNichols (2000) argues that this approach is superior to other accruals models and has the potential for the development of tests which are well grounded in the institutional specifics of accruals thereby providing greater insight into how accruals are managed.

A more recent contribution to the literature on the detection of earnings management comes from Stubben (2010) who develops McNichols argument that a specific accruals method has greater potential for identifying earnings management than aggregate accruals methods. He builds on the evidence that revenue manipulation represents the most common form of earnings management (Dechow and Schrand, 2004; Nelson, Elliott and Tarpley, 2003) and develops a model for detecting earnings management based on the identification of discretionary revenue. The manipulation of revenue and receivables is not specific to one industry and so has the potential not only for greater precision in the identification of earnings management but also for generating findings characterised by a higher degree of generalisability. Stubben's model estimated a firm's discretionary revenues as the residual from the following equation:

$$\Delta AR_{it} = \alpha_1 + \alpha_2 \Delta R_{it} + \varepsilon_{it}$$

Where: ΔAR_{it} is the change in the accounts receivable of firm i between time t and time $t-1$; ΔR is the change in revenue and ε is the residual.

Stubben applied this model to a sample of firms subject to SEC enforcement actions for a mix of revenue and expense manipulation and found that it detected manipulation when accruals models did not. Similar results were found when he simulated manipulation. These findings, he concludes, indicate the revenue model is less biased and better specified than accruals models and argues that estimates from revenue models could be useful as a measure of revenue management and, as a proxy for earnings management, particularly given the predominance of revenue adjustments arising from SEC enforcements.

3.5.4 Manipulation of financial performance: the consequences

As evidenced by empirical literature, the manipulation of reported financial performance is conventionally regarded as leading to adverse consequences: stakeholders are misled about the economic performance of the entity thus leading, by implication, to misallocation of resources; management avoid the consequences of performance measurement or regulatory constraints and, in the case of the manipulation of real transactions, the long term strategic objectives of the organisation may be undermined through lack of investment or inefficient use of resources.

This negative perception is reflected in the following two definitions of earnings management which are taken from two early literature reviews. Schipper (1989) defines earnings management as 'the purposeful intervention in the external financial reporting process, with the intent of obtaining some private gain (as opposed to, say, merely facilitating the neutral operation of the process)'. Healy and Wahlen (1999) extend this view by defining earnings management as occurring 'when managers use judgement in financial reporting and in structuring transactions to alter financial reports to either mislead some stakeholders about the underlying performance of the company or to influence contractual outcomes that depend on reported financial performance'. Dechow and Schrand (2004) also describe accruals

based earnings management, although not necessarily contrary to GAAP, as opportunistically shifting income from one period to another.

However, although a large empirical literature exists which identifies the incidence of earnings management and the incentives which motivate it, few empirical studies address the consequences of eamings management. Of this limited literature most studies are focused on resource allocation decisions of the external stakeholders of public companies. A number of studies, for example, find that the market reaction to the disclosure of earnings management is significantly negative (Foster, 1979; Dechow, Sloan and Sweeney,1996; Beneish, 1997; Palmrose, Richardson and Scholz, 2004). Teoh, Wong and Rao (1998) also find that eamings management contributes to IPO mispricing.

The question of the consequences of earnings management for internal resource allocation is however less well explored. Analytically this question has been investigated by Bar-Gill and Bebchuk (2003), cited in McNichols and Stubben, (2008), who argue that earnings management will lead to inefficient investment because firms overstating their financial results will be able to obtain cheaper financing. Evidence consistent with this hypothesis is found in a recent empirical study by McNichols and Stubben (2008) who, using a number of measures of both earnings management and excess capital investment, find substantial overinvestment in firms engaging in earnings management. The capital investment decisions of three groups of firms over a 25 year period were investigated: those investigated by the SEC for accounting irregularities, those which were sued by their shareholders for improper accounting and those which restated their financial statements. Further, as an alternative proxy for earnings management, discretionary revenues (Stubben, 2010) were also used to identify firms which manipulated earnings. McNichol's and Stubben's findings of inefficient investment in firms manipulating reported financial performance indicate that earnings management, which is largely viewed as influencing the resource allocation decisions of external stakeholders, can also affect internal resource allocation by permitting expenditure levels to be maintained at levels that would otherwise not be possible.

Although the empirical literature adopts a largely negative view of earnings management, a more neutral perspective is adopted in the analytical literature. Scott (2009) argues for example that manipulation which serves to smooth earnings

streams can communicate important decision-relevant inside information to investors and thus have positive consequences. Frequently, agents have specialised information which is prohibitively costly to communicate to the principal (Demski and Sappington, 1987) and manipulation can be used as a means of signalling this information (Demski and Sappington, 1990). However there are also costs associated with earnings management. A perception that reported income is unreliable for resource allocation purposes will give rise to a higher cost of capital and reduced profits, which affect both the investor, through lower stock values and the manager, through lower compensation. Stocken and Verrechia (2004) show that the benefits of earnings management exceed these costs when the amount of inside information is high and a firm's environment is volatile, such that the benefits accruing to insider information are also high. Demski, Frimor and Sappington (2004) also show that principals may benefit from the manipulation of an accounting system where managerial incentives to engage in more damaging responses to the monitoring of financial performance, such as could occur with the manipulation of real transactions, are reduced. However, for benefits to be realised, accounting flexibility should be made available only to skilled managers who will use the flexibility in reporting financial performance to work hard on the principal's behalf. This finding is also consistent with Demski (1998) who formally uses accounting regulation to ensure that income smoothing requires skill and hard work. Thus, Arya, Glover and Sunder (2003) argue that it is the implicit role of regulators to make income smoothing challenging but not impossible. They further argue that the conventional perceptions of income manipulation and transparency are simplistic in that they fail to recognise the ways in which communication is effected across today's large multi-national, multi-layered corporations where total transparency is neither possible nor desirable. Instead of trying to eliminate income manipulation therefore they suggest that it might be more useful to focus on the accounting properties that increase the value of managed accruals. They conclude that the appropriate focus for policy debate is the optimal mix of flexibility and rigidity in accounting regulation and that the current push for increased transparency in corporate financial reporting may be detrimental to shareholder interests.

Thus some flexibility in an accounting system may be beneficial. This is further supported from a contractual perspective as the potential for earnings management may protect investors from the consequences of unforeseen events

when contracts are rigid and incomplete (Scott, 2009). These benefits may be impaired however when the potential for earnings management is too great and the incentive to exert effort is reduced. Scott (2009) thus argues that GAAP, which typically allows some flexibility in the choice of accounting policy, has a positive role to play in allowing some managerial flexibility in the determination of earnings whilst simultaneously restricting the scope for managers to shirk.

Empirical evidence of the impact of tighter accounting regulation aimed at greater transparency in financial reporting and the minimisation of manipulation is limited. However, the passage of the Sarbanes-Oxley Act in 2002 in the US, offers an opportunity for such investigation. A number of studies find significant declines in both total accruals and discretionary accruals in the period following the passage of the Act. (See for example: Jain and Rezaee, 2006; Li, Pincus and Rego, 2008 and Iliev, 2010). However, the impact of this reduced accruals based manipulation is shown to be associated with an increased incidence of transaction based manipulation (Cohen, Dey and Lys, 2008). Whilst acknowledging that increased investor and auditor vigilance may be contributory factors, the authors argue that, faced with reduced scope for manipulating accruals, these findings are consistent with managers seeking alternative, less easily identifiable, ways of manipulating reported financial performance in order to meet financial objectives.

3.5.5 Summary

In summary the analytical literature argues that there are benefits to both managers and principals from a certain level of income manipulation but that these benefits can be eroded if the potential for manipulation is too high. The realisation of the benefits of manipulation can, however, be facilitated by a strong framework of accounting regulation which permits some flexibility in the management of reported financial performance whilst restricting the potential for opportunistic managerial behaviour. The empirical literature in contrast focuses on the negative aspects of manipulation and has identified such manipulation in both the public and private sectors in response to a number of incentives including those relating to accountability, the regulatory environment and the incidence of political costs. The methods adopted to identify income manipulation have however been the subject of some debate. The manipulation of real transactions has been identified in both the public and private sectors in response to target income levels and, in the healthcare

sector, focuses on the potentially adverse consequences of such manipulation for patient care. Accruals based manipulation is perceived as leading to less damaging consequences than the manipulation of real transactions but the number of studies investigating the consequences of accruals based manipulation is limited and largely restricted to the resource allocation decisions of external stakeholders. One recent study however (McNichols and Stubben, 2008) investigates the impact of earnings management on internal resource allocation and finds that it leads to inefficient capital investment decisions in US quoted companies. There are no public sector studies which investigate the resource allocation consequences of the manipulation of reported financial performance.

The identification of manipulation, whether transactions or accruals based, is problematic because managerial motives are not observable. The analysis of earnings distributions provides a method of capturing the combined impact of both types of manipulation but these techniques have been subject to considerable challenge from alternative explanations for observed discontinuities. However, a number of models have been developed for detecting accruals based income manipulation, and, as a consequence, studies into accruals based manipulation now dominate the literature. Aggregate accruals models have, however, also been subject to sustained criticism. As a result, the specific accruals approach, which is associated with greater precision in the measurement of discretionary accruals, is the one most favoured. Empirical investigations using this approach have, however, been largely restricted to the banking and insurance industries. This narrow focus may be a consequence of the somewhat demanding conditions for the application of this model which not only requires the identification of incentives to manage earnings but also needs to be grounded in a deep understanding of the institutional setting and the identification of material specific accruals which are exploitable for earnings management purposes. A recent development in the specific accruals literature, however, which does not suffer from this lack of generalisability, is the development of a model for identifying discretionary revenue (Stubben, 2010). This model is of particular interest as the manipulation of revenue has been found to be a major source of earnings management. No studies in the healthcare literature have, to date, used a specific accruals approach, revenue based or otherwise, to investigate the manipulation of reported financial performance.

The traditional dominant body of earnings management literature focuses only on the consolidated financial statements of quoted companies on US stock exchanges. Elsewhere, however, there is a strand of literature on income shifting between subsidiary companies with the objective of reducing the group's overall burden of taxation. This income shifting is effected through such mechanisms as the location of interest payments and through transfer pricing of goods and services between group entities. Further interesting studies reveal transaction based manipulation of reported financial performance in response to incentives which are specific to institutional settings which differ from the Anglo-American capital markets dominated model of corporate governance and financial reporting. One particular study, set in the Finnish public sector, (Vinnari and Nasi, 2008) reveals income manipulation involving transfers of funds between a City Council and a quasi-independent utility company in order to facilitate the satisfaction of the City Council's statutory breakeven requirement.

3.6 Accruals Accounting and the Application of GAAP in the Public Sector

3.6.1 History of accruals accounting in NHS hospitals

Accruals based earnings management exploits flexibilities in GAAP in order to manipulate reported financial performance. The application of accruals based accounting is, however, a very recent NPM phenomenon in Government departments, where its adoption occurred only the early 21st century. Traditionally public sector accounting has adopted a cash budgeting approach, with little scope for the manipulation of reported financial performance except through the manipulation of real transactions (Lapsley, 1993). However NHS hospital Trusts, consistent with the NPM principles which informed their establishment, have been required to prepare true and fair accruals based financial statements, since their inception in 1991 (NHS and Social Care Act, 1990[12]; HM Treasury Financial Reporting Manual[13] and Department of Health, NHS Finance Manual[14]). Further,

[12] National Health Service and Community Care Act (1990) available at: http://www.opsi.gov.uk/ACTS/acts1990/Ukpga_19900019_en_1.htm

[13] HM Treasury, Financial Reporting Manual available at:

and contrary to common perceptions, NHS hospitals have a long history of accruals based accounting arising from their origins, before the establishment of the NHS, in the private sector (Robson, 2003). Prior to the establishment of the NHS, hospitals operated within the voluntary sector or were run by municipal authorities. As early as 1893, and throughout the first half of the twentieth century, accounting information in hospitals was guided by Burdett's *'Uniform System of Accounts for Hospitals, Charities, Missions and Public Institutions'*. This 'manual' required the production of an income and expenditure account, balance sheet, and invested property account[15]. Further, driven by the needs of key stakeholders, there also developed a tradition of performance measurement, as, for example, in the benchmarking of cost per bed information, an early attempt at measuring the cost efficiency of hospitals. Adopting a social forces model, Jones and Mellett (2007) also argue that hospital and NHS accounting technologies were better developed than commonly perceived. They chart the development of accounting in hospitals according to three periods: pre-NHS, when the emphasis was on stewardship; 1948 -1990 when the emphasis was on budgetary control, not only for the hospital as a whole but also within individual hospital departments; and 1991 onwards, following the introduction of the internal market and the establishment of NHS Trusts, when the emphasis has been on efficiency and performance.

Ellwood (2003) further elaborates on the NHS period prior to the establishment of NHS Trusts when, even whilst being managed as cost centres within local health authorities, a partial accruals approach was adopted. Adjustments were made within the financial statements for receivables, payables and movements in inventory and a statement of financial position which showed current assets and liabilities. Importantly, however, records for property, plant and equipment were isolated and not included in the 'statement of balances'. The exclusion of non-current assets from the financial statements led to accusations of

---

http://www.hmtreasury.gov.uk/frem_previous_manuals.htm

[14] Department of Health, NHS Finance Manual available at: http://www.info.doh.gov.uk/doh/finman.nsf/ManualDownload?OpenView&Start=1&Count=30&Expand=10#10

[15] The income and expenditure account however did not incorporate depreciation or adjustments for inventory.

inefficient use of capital stock and these concerns represented one of the principal arguments for the introduction of accruals accounting throughout the public sector in the early 21st century (Likierman, 1998a,b).

As in the private sector, compliance with GAAP underpins the public sector concept of true and fair. However, adaptations are made where private sector GAAP is considered unsuitable for producing financial statements relevant to the needs of the public sector (HM Treasury, Financial Reporting Manual, Chapter 2, para. 2.1[16]). However, despite the long history of accruals based accounting technologies in NHS hospitals, the adoption and adaptation of private sector GAAP for public entities, including English hospital Trusts, has been subject to considerable critical comment and criticism. This is considered in the following section.

3.6.2 The adoption and adaptation of private sector GAAP

In a review of GAAP application in the public sector, Ellwood (2003), uses examples from the Ministry of Defence, a local authority and an NHS Trust to investigate the extent to which UK public sector accounting is consistent with GAAP. Using illustrations of asset, liability and income recognition, including PFI (Private Finance Initiative) capital projects, she concludes that, as a consequence of these adaptations, the public sector interpretation of 'true and fair' differs from the private sector and further, that adaptations to GAAP vary across these different types of public organisation, thus impairing comparability. She also argues that the public accountability of public sector entities cannot be delivered by private sector GAAP as it was developed with a focus on shareholders as the pre-eminent stakeholder, a notion that cannot readily be transferred to public sector entities.

The transferability of the 'shareholder' perspective of GAAP and adaptations to fit the public sector are the subject of concerns elsewhere in the literature, particularly where adaptations are perceived to have the potential to mislead users. (Hodges and Mellett, 2004, 2005; Barton, 1999, 2000, 2004, 2005; Newberry and Pallot, 2003). In an Australian setting Barton (2005) reviews adaptations to

[16] HM Treasury, Financial Reporting Manual available at: http://www.hmtreasury.gov.uk/frem_previous_manuals.htm

Australian GAAP aimed at improving the efficiency and effectiveness of Australian public sector entities and concludes that these adaptations were inadequate to the task, such that financial statements can be difficult to interpret and do not always report relevant information. Elsewhere, a number of studies utilise the vehicle of PFI projects to illustrate the potential impact of public sector variations from GAAP on the comparability of financial statements, on service provision and on accountability to Parliament. In the UK, Hodges and Mellett (2004) concluded that PFI accounting, which follows a Treasury negotiated variation from GAAP, impaired the comparability of NHS Trust financial statements and the evaluation of Trust performance. Further, they argue that, as a consequence of NHS Trusts' statutory duty to breakeven, the different revenue effects of PFI contracts as compared with hospitals financed by public capital have consequences for service delivery. This concern that the greater revenue costs associated with PFI projects could lead to the reduction of expenditure elsewhere, a form of manipulation aimed at fulfilling accountability requirements and financial objectives, is also echoed by Ellwood, (2008), in a case study review of GAAP modifications in the financial statements of an English hospital Trust. Further, concern that variations from GAAP may impair accountability to Parliament by, for example, allowing the debt associated with PFI schemes to avoid parliamentary scrutiny, has also been expressed by Edwards and Shaoul, (2003) and Newberry and Pallot (2003).

Elsewhere the process of determining public sector variations from GAAP, focusing in particular on accounting for PFI schemes, has been variously interpreted from a sociological perspective by Broadbent and Laughlin (2005) using a Habermasian framework, by Hodges and Mellett (2005) using the concept of omitted influences; by Rutherford (2003) adopting a social constructionist perspective and by Broadbent and Laughlin (2002) adopting a Marxist perspective. Hodges and Mellet (2002) further use the case study of accounting for PFI projects to gain further insights into Walker and Robinson's (1993) model of accounting standard development.

In summary, the public sector accounting literature on variations from GAAP is dominated on the one hand by critical review and commentary and, on the other, by interpretive studies, which adopt a variety of theoretical perspectives to examine the process of determining and agreeing variations from GAAP. Empirical evidence to support the critical commentary has been provided by case study illustrations (as

in: Ellwood, 2003, 2008; Hodges and Mellett, 2004), interview based studies as in Hodges and Mellett (2005) and contextual analysis (as in Broadbent and Laughlin, 2002). The impact of variations from GAAP on both the performance and accountability of public sector entities has featured as an issue worthy of analysis within the wider context of their impact on public sector financial reporting but has yet to be the subject of empirical archival investigation.

3.7 SUMMARY AND CONTRIBUTION

Over the last thirty years the UK public sector has been subject to a wave of reforms which has collectively become known as New Public Management. New Public Management has been informed by a belief in market based solutions, drawing on private sector models of management and governance to address concerns about the performance, efficiency and accountability of public sector entities. NPM initiatives, such as the separation of the purchasers and providers of hospital services in a quasi-market relationship have, since the 1990s, served to bring greater emphasis to agency relationships in an environment previously characterised by hierarchical modes of governance. In the 21st century further NPM initiatives have followed, aimed at the mitigation of agency costs and the improvement of both quality and efficiency a quasi-market environment. These initiatives have included the introduction of systems of performance measurement and management in the form of the Star ratings and the Annual Health Check, and the development of a costing system, in the form of the reference costing system, which has facilitated both the development of a uniform pricing system for hospital services and the benchmarking of Trust cost efficiency. As a means of enhancing the accountability of public sector organisations much of the data relating to these initiatives has been made publicly available. To date, however, this data remains largely unexplored.

Concerns over the cost efficiency of NHS hospitals have persisted almost since the NHS' inception in 1948 and have been a key driver of NHS reforms throughout its history. However, inadequate investment in costing systems impeded the operation of the internal market and the benchmarking of Trust cost efficiency until the early 21st century, when the new Labour Government made a commitment to the development of the reference costing system. To date, public sector accounting research has focused on an evaluation of the reference costing system

in isolation of other NPM initiatives and has adopted largely critical and case study approaches. However, despite being subject to early criticism, reference costs are now used by a wide range of users for a wide range of purposes, including contracting between purchases and providers, and the UK is regarded as being in the vanguard of cost benchmarking for hospitals.

The systematic collection of reference costs with adjustments for case mix and market factors render them a more reliable measure of cost efficiency than measures of cost efficiency obtained using frontier analysis techniques, such as data envelopment analysis and stochastic frontier analysis. Despite a very large literature investigating hospital efficiency, these techniques have been subject to considerable criticism because of the sensitivity of findings to model specification, choice of input and output variables and a failure to explain the relationship between variables. Further, the issue of the quality of hospital services in these studies is problematic. If quality is a hidden component of costs then the findings from such studies could lead to inappropriate policy initiatives which, in the pursuit of efficiency, serve to reduce the quality of care. Studies which aim to investigate the relationship between cost efficiency and quality have therefore adopted a two stage approach where measures of cost efficiency are obtained using frontier analysis techniques which are then used in a regression with measures of quality.

However the investigation of the relationship between quality and efficiency has been further impeded by the challenge of measuring quality. Measures of quality such as mortality and readmission rates are both narrow and problematic in that higher rates are not necessarily indicative of poor quality. However broader measures, which necessarily involve bringing together many aspects of care across many individual hospitals, have tended to rely on staff and patient perceptions obtained using survey instruments. The introduction of multi-dimensional performance measurement in the English NHS applied uniformly to all hospital Trusts offers an opportunity to develop a more objective and broader measure of nonfinancial performance than has hitherto been possible.

The introduction of multi-dimensional performance systems in the NHS has been the subject of a number of studies in the public sector accounting literature. These have largely adopted a critical review or interpretive method and have focused on the tensions between different modes of governance, the political

context in which performance measurement was introduced and the potentially adverse consequences for patient care. This last theme is mirrored in a number of empirical archival studies which can be found in the economics literature where the focus is mainly on the dysfunctional responses to performance measurement and the investigation of trade-offs, manipulation and gaming. However, evidence from these studies is consistent with a story of performance targets incentivising process improvements which have beneficial effects for service standards and quality of care throughout the hospital.

The potential for manipulation of performance measurement systems comprises a significant strand of the performance measurement literature in both the public and private sectors. Financial measures of performance retain a key position in these systems and the manipulation of financial performance measures has received extensive attention in the literature. Over the last 20 years however this literature has become dominated by the 'earnings management' literature which is largely concerned with the manipulation of financial performance in US quoted companies. This literature has identified earnings management in response to a number of incentives of which some are specific to a capital markets context whilst others, such as regulatory and political incentives, are common to many institutional settings, including the public sector. Elsewhere, however, there are other streams of literature which often refer to income smoothing, income shifting or creative accounting. These studies include the investigation of the manipulation of financial performance in individual group entities, rather than at corporate level, and in institutional settings which differ from the conventional Anglo American corporate model. The income shifting literature, for example, investigates the manipulation of subsidiary financial performance through the structuring of transactions between group entities, the ultimate objective of which is an economic benefit in the form of a reduction in the overall group tax burden. A small number of studies also investigates the manipulation of reported financial performance in different institutional settings. One study in particular, (Vinarri and Nasi, 2008) investigates income shifting between a Finnish City Council and its 'wholly owned' water utility company in order to facilitate the satisfaction of the City Council's statutory requirement to breakeven. These studies, whilst set in a different institutional context, lend an additional perspective to McNichols (2000) argument that the development of tests which are well grounded in the institutional specifics of

accruals have the potential to provide greater insight into earnings management than the application of conventional accruals based models and distributional analysis.

The empirical literature on the manipulation of financial performance adopts a largely negative perspective arguing that such manipulation undermines managerial accountability and leads to inefficient resource allocation. The analytical literature, in contrast, argues that some flexibility in accounting regulation, which permits manipulation of reported financial performance, can generate benefits to principals. The facility for smoothing income however must be restricted to 'good' managers who use the flexibility to work hard in the interests of principals. A strong framework of accounting regulation which permits some flexibility in the reporting of financial performance whilst restricting the potential for poorer quality managers opportunistically to exploit this flexibility is seen as a key mechanism for delivering such benefits. In contrast, too much rigidity in the interests of transparency may perversely lead to adverse consequences. For example, managers may resort to other tactics, which are more damaging to the long term strategic capacity of the organisation, in order to meet financial objectives.

In the public sector a notable NPM development, aimed at improving accountability, has been the adoption of accruals accounting based on generally accepted accounting principles (GAAP). A considerable literature reviews the merits of the adoption of GAAP and its adaptation to fit the public sector setting. Most of this literature employs a critical review method. No papers have yet empirically investigated the performance impact of variations from GAAP.

This thesis comprises two main studies: Study 1 investigates firstly, the relationship between cost efficiency and service standards in NHS hospitals over the period 2003-2008 and secondly, the differential impact on this relationship when the Annual Health Check replaced the Star ratings in 2006; Study 2 investigates the impact on the performance and accountability of NHS Trusts of an NHS specific transaction, known as financial support which operated so as to transfer funds into Trusts in financial difficulty from surplus funds elsewhere in the NHS. The accounting treatment of financial support, in an approved variation from GAAP, permitted flexibility in reported financial performance, often by advancing the

recognition of revenue, with the objective of maintaining service standards at a higher level than would be possible within the breakeven constraint.

This thesis makes the following contributions to the literature. First, it makes a contribution to the limited number of empirical archival studies in public sector accounting research in the UK, Europe and Australasia, particularly in the domain of performance measurement where a qualitative paradigm is dominant. Secondly, it makes a contribution to the literature on cost efficiency in hospitals by investigating the relationship between service standards and cost efficiency in NHS hospitals. Thirdly, it contributes to the literature on public sector performance measurement, including the balanced scorecard literature, by investigating, in a quasi-experimental setting, the relative impact of performance measurement system design on the incentivisation of cost efficiency in English hospitals. Fourthly, it makes a contribution to the public sector literature on the manipulation of financial performance, and to the wider literature on the manipulation of financial performance, by investigating whether a form of income manipulation, as manifest by the accounting policy for financial support in English acute hospitals, can generate benefits to stakeholders in the form of better performance. Fifthly, it adds to the literature on the manipulation of financial performance and contributes to our understanding of such manipulation, by conducting an investigation in a novel setting which is characterised by a complex multi-layering of incentives interacting in an institutional context which has been subject to a sustained programme of reform. Lastly, it makes a contribution to the literature on accountability and the application of UK GAAP in the UK public sector by evaluating the impact of the policy of financial support on NHS Trust accountability to Parliament and the public.

CHAPTER 4 INSTITUTIONAL CONTEXT

4.1 INTRODUCTION

NHS Trusts were established in 1991 as a significant policy innovation aimed at improving the efficiency of public sector hospitals. Although subject to regulation by the Department of Health, NHS Trusts are publicly funded, quasi-independent organisations which mimic private sector organisations in their governance arrangements (National Health Service and Community Care Act, 1990[17]). They generally comprise one or more hospitals whose revenue is derived mainly from contractual relationships with healthcare commissioners, Primary Care Trusts (PCTs), which hold the majority of the NHS budget allocation. NHS Trusts generally contract with ten or fewer PCTs although some large Trusts, embracing more specialist services, may provide services more widely. PCTs commission acute hospital services for their local health economies through annually negotiated service level agreements mainly, though not exclusively[18], with NHS Trusts.

NHS Trusts thus represent a stereo-typical NPM type reform (Hood 1991, 1995). They were disaggregated from the Department of Health as corporatised units; their relationships with commissioners became contractualised rather than being hierarchical and their governance was effected by a unitary board modelled on best practice in the private sector (Combined Code, 2003[19]; Financial Reporting Council, 2006). As a measure to enhance the public accountability of hospitals each NHS Trust was subject to a statutory duty to breakeven, satisfaction of which was made by reference to their financial statements. In contrast with Government departments at the time, NHS Trusts were from their inception in 1992, required to

[17] National Health Service and Community Care Act (1990) available at: http://www.opsi.gov.uk/ACTS/acts1990/Ukpga_19900019_en_1.htm

[18] Primary Care Trusts are expected to commission services for patients from a diverse range of providers which include not only NHS Trusts but also Foundation Trusts, (a new form of public benefit entity), independent sector treatment centres, private hospitals and hospitals overseas. The commitment to enhanced diversity and plurality of provision was set out in 'Delivering the NHS Plan' (Department of Health, 2002a). Thus NHS Trusts compete with a variety of providers to deliver services.

[19] Combined Code (2003): '*Combined Code on Corporate Governance*' available at: http://www.fsa.gov.uk/pubs/ukla/lr_comcode2003.pdf

prepare GAAP compliant[20] accruals based financial statements. This requirement was not imposed on Government departments until the early 21st century.

However, towards the end of the 1990s there was considerable public concern about the performance of the NHS and, following a brief respite in the programme of NHS reforms immediately after the election of the Labour Government in 1997, the NPM reform agenda continued to be driven forward. Consistent with a principal-agent perspective (Jensen and Meckling, 1976) whereby agents incur monitoring costs to reduce information asymmetry and to align incentives with principal's objectives, the Government, in the form of the Department of Health, acting as agent to Parliament and the public, introduced a number of significant reforms aimed at incentivising cost efficiency as a means of delivering high service standards. These included the development of a system for measuring Trust cost efficiency and, to ensure that cost efficiency was not pursued at the expense of quality, the implementation of a multidimensional performance measurement system (PMS). The Star ratings system was introduced in 2001-02 and was replaced by the Annual Health Check (AHC) in 2006. Common features of these two PMS facilitate an investigation into the relationship between cost efficiency and quality of services in NHS hospitals, a key concern of the literature on hospital cost efficiency, during the period 2003-2008. Further, the common characteristics of each PMS facilitate an investigation into their relative effectiveness at incentivising cost efficiency.

A key performance indicator in both the Star ratings and the AHC was reported financial performance. In the Star ratings, Bevan (2006) identified the potential for enhancing a Trust's performance rating by breaching the requirement to breakeven and incurring a deficit in order to improve performance against nonfinancial indicators. Deficits in the NHS could however be disguised by accessing the system of financial support, which served to manipulate reported financial performance by shifting income between NHS organisations and accelerating revenue recognition in Trusts in financial difficulty. The system of

[20] GAAP compliance was determined by reference to private sector GAAP as interpreted in the NHS Accounting Manual as being appropriate and relevant to the NHS. The NHS Accounting Manual can be found at: http://www.info.doh.gov.uk/doh/finman.nsf/ManualDownload?OpenView

financial support was deeply embedded in the institutional structure and history of the NHS and allowed a degree of flexibility in reported financial performance in order to meet accountability requirements, most notably, the statutory duty to breakeven. The operation of the system has, however, been largely hidden from public scrutiny.

The main objectives of this thesis are to investigate the performance and accountability consequences of two aspects of performance measurement. The first study investigates the relationship between cost efficiency and nonfinancial performance over the period 2003-2008 and the relative effectiveness of the Star ratings as compared with the AHC in incentivising cost efficiency as a means of improving service standards. The second study focuses on the manipulation of financial performance through the application of the system financial support and investigates the accountability and performance consequences of such a system. The main objective of this second study is to evaluate whether financial support, which accorded some accounting flexibility in the reporting of financial breakeven, generated benefits in the form of better overall performance in those Trusts which received it.

This Chapter of the thesis sets out the institutional framework which provides the setting for these two investigations. Section 4.2 reviews the financial accountability of NHS Trusts; Section 4.3 considers the issue of cost efficiency and the development of its measurement; Section 4.4 reviews the two performance measurement systems which were introduced with the objective of incentivising the delivery of cost effective services; Section 4.5 considers the NHS specific transaction 'financial support' and its impact on Trust performance and accountability; Section 4.6 concludes with a summary of the institutional issues relevant to the investigations in this thesis.

4.2 THE FINANCIAL ACCOUNTABILITY OF NHS TRUSTS

Despite notionally operating outside the boundaries of the Department of Health, NHS Trusts are financed wholly by public capital, on which an annual dividend (capital charge) is paid, and are accountable to the Department of Health for their performance. However, in contrast with private sector wholly owned subsidiaries, with which they might be compared, the reported financial performance

of individual NHS Trusts was accorded particular significance in the satisfaction of public accountability. First, Trusts have a statutory duty to breakeven and satisfaction of this duty is determined by reference to reported financial performance (National Health Service and Community Care Act, 1990). Secondly, the number of Trusts in deficit and in breach of their statutory duty is reported to Parliament in the annual NHS Summarised Accounts[21], prepared by the Department of Health.

4.2.1 The statutory duty to break-even

The statutory duty to break even is laid out in the National Health Service and Community Care Act, 1990, Section 10[22] and requires a Trust 'to ensure that its revenue is not less than sufficient, taking one financial year with another, to meet outgoings properly chargeable to the revenue account[23]'. NHS Trusts normally plan to meet their statutory duty by achieving a balanced position on their income and expenditure account in each and every year, as required by the Department of Health, but the phrase 'taking one financial year with another' does provide a degree of flexibility about the timescale for matching income with those costs whose incidence is uneven, such as the costs associated with the early retirement of staff, and when managing the recovery of an NHS Trust with serious financial difficulties. The breakeven duty is therefore assumed to have been met if a material cumulative deficit position on the income and expenditure account is recovered over a three year period (NHS Executive, 1999). A deficit is regarded as material if it exceeds more than 0.5% of total annual turnover. Exceptionally, and with the express agreement of the relevant SHA, the recovery period can be extended to five years.

4.2.2 The wider financial accountability of the NHS to Parliament and the public

A key motivation for the introduction of NHS Trusts was the improvement of public accountability for hospital performance both to Parliament as a whole and to

---

[21] See NHS Summarised Accounts from 2002-03 available at: http://www.nao.org.uk/publications/0304/nhs_england_summarised_accou.aspx

[22] National Health Service and Community Care Act 1990, s.10 available at: http://www.legislation.gov.uk/ukpga/1990/19/section/10

[23] In this context the 'revenue account' refers to the income statement.

local communities. The financial accountability of NHS Trusts is discharged through the submission to the Department of Health of annual audited financial statements under the provisions of the National Health Service Act, 1977, Section 98(2)[24]. These financial statements are prepared in accordance with the requirements of the NHS Accounting Manual and form the basis for the preparation of the Summarised Accounts of NHS Trusts[25]. These accounts are laid before Parliament, as principal, by the Permanent Secretary of the Department of Health, (who is also the CEO of the NHS), in his role as agent.

The NHS Accounting Manual[26] is prepared in accordance with the Treasury's Financial Reporting Manual[27] (FReM), the contents of which are determined after consultation with the independent Financial Reporting Advisory Board (FRAB). The accounting policies contained in the FReM follow UK generally accepted accounting practice for companies (UK GAAP) 'to the extent that it is meaningful and appropriate to the public sector'.

A further element of the public accountability of NHS Trusts is the requirement to present Trust financial statements at an annual public meeting (modelled on the private sector annual general meeting). However, attendance at these meetings is low, with the number of Trust employees often exceeding the number of members of the public (Hodges, MacMiven, Mellett, 2004) and unless members of the public attend this meeting they generally do not have access to the full audited financial statements which are remitted to the Department of Health. Instead, summarised accounts, containing much less information, are made available in the Trust's Annual Report which is generally made more widely available, often both electronically on the Trust's website and in hard copy.

The financial accountability of the Department of Health, to whom NHS Trusts report, is discharged, as for other Government Departments, through the

---

[24] Available at: http://www.legislation.gov.uk/ukpga/1977/49/contents

[25] Available at: http://www.official-documents.gov.uk/

[26] Available at: http://www.info.doh.gov.uk/doh/finman.nsf/ManualDownload?OpenView

[27] Available at: http://www.hm-treasury.gov.uk/frem_previous_manuals.htm

submission of annual audited true and fair financial statements to Parliament under the provisions of Section 7 of the Government Resources and Accounts Act, 2000[28]. These financial statements, submitted to Parliament in addition to the Summarised Accounts of the NHS, are prepared in accordance with the requirements of HM Treasury's FReM and consolidate the financial information of bodies operating within the Departmental Accounting Boundary, (see Figure 3), most notably PCTs (which are budget holders for the majority of the Department's budget), SHAs and the administrative functions of the Department. The Departmental accounting boundary is different from the concept of the group in the commercial sector as it is based on in-year budgetary control, with Departments being required to live within an allocated budget[29]. NHS Trusts, as provider organisations, lie outside the Departmental Accounting Boundary, gaining their revenue not from budget allocations but 'direct from trading activities' (Department of Health Resource Accounts, 2006-07). However the surpluses and deficits of NHS Trusts do contribute to a determination of whether the Department has lived within its allocated resources (the Departmental Expenditure Limit). The Permanent Secretary, in addition to being Accounting Officer for the Department of Health is, therefore, also accountable to Parliament for expenditure in NHS Trusts (National Health Service Act 1977, s 98(4)), through the submission of 'Summarised Accounts of the NHS'. In a reflection of the significance of the statutory duty to breakeven for the public accountability of NHS Trusts, these financial statements include reports of the number, but not the identity, of Trusts failing to breakeven, even if, taken together, all NHS Trusts achieve financial balance.

[28] Government Resources and Accounts Act, 2000 available at: http://www.legislation.gov.uk/ukpga/2000/20/contents
Department of Health resource accounts can be obtained from the Department of Health as at:
http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_081317

[29] See Department of Health Resource Accounts 2006-07 available at:
http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_081317

4.2.3 Further significance of financial break-even for public accountability

In addition to having a key role in the discharge of the public accountability of NHS Trusts, the achievement of financial breakeven has had further significance for the accountability of NHS Trusts as it has been conventionally regarded as a proxy measure for Trust cost efficiency. This issue is considered in detail in Section 4.3 below. It has also played a significant role in determining overall Trust performance ratings following the introduction of a multi-dimensional performance measurement system in 2001-02. This issue is considered in Section 4.4.

4.3 PERFORMANCE MEASUREMENT: COST EFFICIENCY

During the 1980s there was considerable investment in the development of costing systems (Ellwood 1996) but, despite this investment, there was very little information at the level required to form a basis for cost based pricing between purchasers and providers, an integral part of the vision for the internal market (Ellwood, 1995). As a result a significant proportion of the contracting between purchasers and providers was effected by means of block contracts, where broadly based costing information based on historical allocations could be used as the basis for negotiation. Further, government measures of cost efficiency during this period were subject to considerable criticism (Jacobs, 2001; Street and Jacobs, 2002; Street 2003). In this environment, financial breakeven became a proxy for assessing Trust cost efficiency. Trusts breaking even were conventionally regarded as being efficient and those which incurred a deficit, as being inefficient. In the absence of effective costing systems however concerns about variations in performance which were considered unacceptable (Department of Health, 2000) persisted. In response, the NHS Plan (Department of Health, 2000) made an explicit commitment to the combined pursuit of quality and efficiency (Department of Health, 2000) and the development of a Performance Assessment Framework (Department of Health, 1999). In this Framework efficiency was one of six key aspects of performance that would deliver 'high quality, cost effective services' (p.8, para 15). New measures of efficiency, most notably Trust reference cost indices, were to be developed to support both the commissioning of services and the benchmarking of performance. Thus, despite longstanding concerns over efficiency in the NHS, a key driver for many NPM type reforms, it was only towards the end of the 20th century that effective measures of Trust efficiency, Trust reference cost indices, were developed.

Reference cost information was first collected in 1997-98. However, early concerns about data quality (Department of Health, 2002b; Audit Commission 2004) may provide a partial explanation for the omission of reference costs from the multi dimensional performance measurement systems, the Star ratings system (2001-02 to 2004-05) and the subsequent Annual Health Check (2005-06 to 2008-09) which were applied to NHS Trusts from 2001. As a consequence, the commitment to include measures relating to cost efficiency (Department of Health, 1999) thus defaulted to the proxy of financial breakeven despite the susceptibility of reported financial performance to manipulation. Substantial investment was subsequently made into improving reference costs. This was driven in part by the demands of a new hospital payment tariff system which required robust reference costs for individual treatments. The introduction of this system, Payment by Results (Department of Health, 2002b), which commenced in 2003-04, was in turn motivated by the need for an equitable system of hospital payment, uniform across all healthcare providers. This was an essential pre-requisite for a key Government policy objective: the opening up of the market for providing NHS hospital services to a greater diversity of providers, including those from the private sector.(Department of Health, 2003a). Consistent with the Government's objectives, this system incentivises the development of a market in quality rather than one based on price competition, which could have adverse consequences for patients.

Under the reference cost system actual costs for individual treatments, calculated on a full cost absorption basis, adjusted for case mix and market factors arising for example from geographical location, are centrally collected by the Department of Health. Trust reference cost indices are weighted aggregate measures of Trust cost efficiency based on the actual costs for individual treatments[30] (Department of Health, 2006a, b). The Reference Cost Index represents the cost of a Trust's aggregate activity compared with the national average cost for the same treatments. A reference cost index of 100 represents a

[30] For reference to the method of calculation see for example http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4133221 and http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_118338.pdf

Trust operating at the average level of cost efficiency, one of 101 represents proportionately higher costs and 99 represents proportionately lower costs. As Trusts based in some areas of the country, suffer higher than average costs for staff and land and buildings because of external market forces, the reference cost index is adjusted by a Market Forces Factor. Where costs are higher than average, as in London and the South East, the Market Forces Factor is greater than 1 and operates to reduce the Reference Cost Index. Where costs are lower than average, as in Cornwall, the Market Forces Factor is less than 1 and operates to increase the Reference Cost Index (Department of Health, 2006 a, b). Further, cost data is truncated by excluding bed days that fall outside of nationally set lengths of stay (trimpoints). The costs of any days beyond these trimpoints are excluded from the analysis. This assists in giving a like-for-like comparison of activity and costs. As a control measure, Trust reference costs are reconciled to the financial statements to ensure that all relevant costs are reported. The weighted aggregate reference cost index for each Trust thus represents an inverse measure of relative cost efficiency. Trust reference cost indices are independent of the both the Star ratings and the AHC, cover most hospital activity and, unlike reported financial performance which is subject to the statutory duty to breakeven, do not form a key part of the performance management framework. They are therefore relatively robust against manipulation and represent a key element in the research design of the two studies in this thesis.

4.4 PERFORMANCE MEASUREMENT: MULTI-DIMENSIONAL SYSTEMS

4.4.1 Introduction

During the Thatcher regime of the 1990s, consistent with a general policy of deregulation, and the adoption of market based regulatory mechanisms, the performance measurement and management of NHS Trusts focused primarily on financial performance and the fulfilment of their statutory duty to breakeven. However, a parallel initiative with the introduction of NHS Trusts in 1991 was the introduction of the `*Patient's Charter'*. The Charter spoke in terms of patients having rights in relation to a limited number of targets mostly related to waiting times for treatment. However although there was some reduction in waiting times and in the way hospitals conducted themselves, the initiative was one of rhetoric as much as

action. (Bristol Enquiry, 2001, Chapter 4 para. 26). By the time the Labour Government was elected in 1997 continued pressure on resources for the NHS had resulted in considerable public concern about access to services and the NHS Plan (Department of Health, 2000) identified a reduction in waiting times as the public's top priority (Department of Health, 2000, Chapter 12). In response the Government, consistent with a command and control hierarchical style of governance, (Le Grand, 2010; Ouchi, 1980) introduced a Performance Assessment Framework (Department of Health, 1999), aimed at securing both cost effective and quality healthcare services through the setting, implementation and monitoring of national standards for treatment. Following the pilot testing of a limited number of performance standards (high level performance indicators, HLPIs) in Health Authorities, a multi-dimensional system of performance measurement, the Star ratings was rolled out into NHS hospital Trusts in 2001. Publication of the first Star ratings occurred for the year ended 2001-02 with a view to enhancing the public accountability of NHS Trusts for their performance (Department of Health, 2002a). This section proceeds with a review of the Star ratings (Section 4.4.2) followed by a review of the Annual Health Check (Section 4.4.3). It concludes with a summary (Section 4.4.4) of the significance of this institutional setting for research investigations.

4.4.2 The Star ratings

The Star ratings system delivered a composite measure of performance, the Star rating, which represented an overall assessment of performance against a number of individual component measures, across a range of perspectives inspired by the balanced scorecard. Schwartz et al, (2011) argued such a balanced scorecard approach, employing both composite measures, which are useful for policy purposes, and individual component measures, which discourage trade-offs, may be most effective at incentivising the delivery of high service standards cost effectively.

There were two key elements in the Star ratings system, determined after extensive consultation with the public and the professions. First, a limited number of key targets was identified as representative of the public's top priorities for hospital services and second, a broader set of measures reflected the wider interests of key stakeholders.

There were nine key measures in each of the years of the Star Ratings (2002-03 to 2004-2005[31]) with the exception of 2005 when there were eight[32]. These measures were linked to Government commitments made in the NHS Plan (Department of Health, 2000) and were dominated by measures related to access to treatment (waiting times). The five key measures which related to waiting times for treatment and were consistent throughout the period of this study were: four hour and twelve hour accident and emergency waits; cancer waits; and inpatient and outpatient waits. Of the remaining four measures, reported financial performance and hospital cleanliness also featured in each year of the Star ratings. The remaining key targets were made up of one or two of the following: an 'improving working lives' measure; a patient appointment booking measure and a cancelled operations measure. In addition to the nine key measures there were approximately thirty other 'balanced scorecard' measures clustered in three perspectives: the clinical perspective, the patient perspective and the capacity and capability perspective. In contrast with the key measures which were very much driven by the priorities of the public as identified by the new Labour Government, these additional measures were determined by the independent performance regulator, the Commission for Health Improvement, and the Department of Health after further consultation with service and other stakeholders. As their collective name suggests, these additional measures were inspired by the notion of the balanced scorecard (Kaplan and Norton, 1992, 1996a). The patient perspective can be seen to be broadly consistent with the customer perspective; the clinical perspective is similar in purpose to the internal business process perspective and the capacity and capability perspective with the learning and growth perspective. There is, however, no financial perspective and no evidence that the causal mapping recommended by Kaplan and Norton (1996a) was undertaken. The patient perspective and the capacity and capability perspective, consistent with the arguments put forward in Ludwig, Van Merode and Groot, 2010 and Schwartz et al., 2011, included not only measures such as waiting times for treatment but also measures of service quality

[31] The Star ratings system was initially introduced in 2001-02 but the system was not fully developed and operational until 2002-03.

[32] The original intention was that there would be nine key targets in each of the years of the Star ratings but in 2005, the Improving Working Lives measure, which was concerned with new ways of working was dropped as it was deemed to have been fully implemented.

based on the results of patient and staff survey data. A full list of the Star rating performance indicators can be found in Appendix 1.

Under the Star rating system, which was devised and regulated by a newly formed regulator, the Commission for Health Improvement, performance against the key targets was measured against a three point categorical scale, according to whether the predetermined target was achieved, underachieved or significantly underachieved. A negative scoring system was operated: zero penalties for achievement, two for underachievement and six for significant underachievement. These scores were aggregated to obtain a total penalty score. Performance against balanced scorecard measures, in contrast was scored positively with higher scores awarded for achievement of the target. These scores were similarly aggregated and then an overall 'balanced scorecard' score of between zero (worst) and six (best) was awarded based on the Trust's performance relative to all other Trusts. These two main performance scores (for the key targets and the balanced scorecard) were then aggregated according to another scoring matrix to determine the final rating of the Trust on a scale of zero stars (worst) to three stars (best). Figure 5 shows the significance given to the key targets by the Government in this aggregation. The limited number of key targets are given a much heavier weighting in the overall rating than the much larger number of balanced scorecard targets. This can be illustrated by identifying that a two star rating can be achieved by concentrating on the nine key targets, where up to two penalty points can be incurred for underachievement of one of the performance standards, with little attention to the much larger number of balanced scorecard targets. As can be seen, a two star rating, which removes the CEO from additional monitoring and threat of dismissal, can be obtained with a score as low as one for the balanced scorecard measures.

Figure 5: Determination of star rating – scoring methodology[33]

	Balanced scorecard* points						
9 key targets – total penalty points	**0**	**1**	**2**	**3**	**4**	**5**	**6**
>12	0★	0★	0★	0★	0★	0★	0★
7-12	1★	1★	1★	1★	1★	1★	1★
3-6	1★	1★	1★	1★	2★	2★	2★
0-2	1★	2★	2★	2★	2★	3★	3★

*c. 30 indicators across 3 perspectives

Source: Performance ratings methodology for 2003/04 (Acute hospital Trusts), Commission for Health Improvement, July 2004.

Trust Star ratings were linked to a number of performance management mechanisms aimed at incentivising good performance as measured by the PMS (Department of Health, 2002c). These included rewards for good performance ratings in the form of lower levels of monitoring, the ability to apply for Foundation Trust status and reputational gains. Penalties for poor performance included the risk of dismissal and reputational damage. The CEOs of zero and one star Trusts were given a limited time of as little as three months to improve performance or face the possibility of dismissal. In a study of CEO incentives in NHS Trusts during this period Ballantine, Forker & Greenwood, (2008) found that CEO turnover was higher than that identified in prior private sector and hospital based-studies.

4.4.2.1 Financial measures in the Star ratings

Within the Star ratings system there was only one financial measure: reported financial performance. This measure acted as a proxy for the cost efficient

[33] A more detailed description of the derivation of the scores on the balanced scorecard and on the key target penalties can be found in Appendix 2. Further information on the performance ratings methodology can also be found at:
http://ratings2005.healthcarecommission.org.uk/methodology.asp
http://ratings2004.healthcarecommission.org.uk/methodology.asp

delivery of services and was measured, as for all key Government measures, on a three point categorical scale: zero penalty points for breakeven or better; two penalty points for a small deficit of up to 1% of revenue and six penalty points for a large deficit of more than 1% of revenue.

4.4.3 The Annual Health Check

On the 1st April 2004, the Commission for Health Improvement, which was established to oversee and manage hospital inspections and performance ratings when the Star ratings were introduced, was replaced by the Healthcare Commission. The Healthcare Commission brought together, under the management of one organisation, several different strands of performance monitoring for the purposes of better co-ordination across the NHS. In addition to the work of the Commission for Health Improvement the Healthcare Commission also took over the work of the National Care Standards Commission, which had regulated NHS services delivered in the private and voluntary sectors, and also assumed responsibility for the value for money work which had previously been the responsibility of the Audit Commission. In 2004-05 the Healthcare Commission continued to operate the Star ratings system which it had inherited from the Commission for Health Improvement. However, the Commission was concerned about growing criticism of the opportunities for manipulation that had been identified in the Star ratings (Bevan and Hood, 2006; Bevan 2006) and the need to integrate new forms of provider, particularly Foundation Trusts. Consistent with the conclusions of Kelman and Friedman (2009) and the findings of Courty and Marshcke (2003b, 2007) that gaming of performance measurement is revealed over time and that managerial responses lead to evolutionary PMS design, the Healthcare Commission introduced a new system, the Annual Health Check in 2005-06 (Healthcare Commission, 2005a).

There were two key differences between the Star ratings and the AHC. First, the AHC was characterised by additional measurement diversity, incorporating approximately 70 nonfinancial measures as compared with less than 40 in the Star ratings and second, the weighting given to financial measures was much higher in the AHC than in the Star ratings. First, the incentive for trading financial for nonfinancial performance was reduced by the non-aggregation of financial measures with nonfinancial measures. Instead of one overall rating the Annual

Health Check had two components: the 'Quality of Services' rating, which captured performance against the nonfinancial measures, and the 'Use of Resources' rating which captured performance against the financial measures. Second, the number of financial measures was increased, with the introduction of five perspectives of financial performance, resulting in fourteen principal measures of performance. Third, for the highest rating, there was a requirement for persistence over a three year period in the achievement of financial breakeven whilst a deficit was heavily penalised, automatically resulting in the lowest overall Use of Resources rating, no matter how good the Trust's performance on other measures.

The two principal elements of the Annual Health Check, the 'Quality of Services' and 'Use of Resources' are considered in more depth in the following sections.

4.4.3.1 Nonfinancial measures in the Annual Health Check: Quality of Services

There were three main components of the Quality of Services rating: Existing National Targets (broadly equivalent to the Key Targets in the Star ratings), Core Standards (broadly equivalent to the Balanced Scorecard) and New National Targets.

Existing National Targets

The twelve Existing National Targets were broadly equivalent to the nine Key Targets in the Star ratings and represented key Government priorities for hospital services. The additional measures: thrombolysis waits, revascularisation waits and rapid access to chest pain clinics, had been present in the Star ratings but as the lower weighted Balanced Scorecard indicators, rather than as Key Targets. A full list of the Existing National Targets can be found in Appendix 3. Performance against the Existing National Targets (ENTs) was, as with the Star ratings, measured on a three point categorical scale but, in contrast with the Star ratings, the scoring was positive: three points for achievement, two for underachievement and zero for failure.[34] The total ENT score then determined a ranking based on a

[34] Scoring methodology for each year of the Annual Health Check can be obtained from: http://www.cqc.org.uk/guidanceforprofessionals/nhstrusts/annualassessments/annualhealthcheck2005/06-2008/09.cfm

four point scale: not met (worst), partly met, almost met and fully met (best). As not all Trusts were subject to all twelve targets the scores were scaled to determine this overall level of performance. Figure 6 illustrates how the final categorical ranking was determined:

Figure 6: The Annual Health Check: Overall Scoring for Existing National Targets

Number of applicable targets	Maximum points available	Overall score			
		Fully met	Almost met	Partly met	Not met
12	*36*	>=33	>=30	>=27	<27
11	*33*	>=30	>=27	>=24	<24
10	*30*	>=27	>=24	>=21	<21
9	*27*	>=25	>=22	>=19	<19
8	*24*	>=22	>=20	>=17	<17
7	*21*	>=19	>=17	>=15	<15
6	*18*	>=17	>=15	>=13	<13
5	*15*	>=14	>=12	>=11	<11
4	*12*	>=11	>=10	>=9	<9
3	*9*	>=9	>=8	>=7	<7

Core standards

The Core Standards were broadly equivalent to the Balanced Scorecard of the Star ratings but the number of perspectives was increased to seven: safety; clinical and cost effectiveness; governance; patient focus; accessible and responsive care; care environment and amenities, and public health. There were 24 measures in total spread across these perspectives but many of these had subsidiary measures resulting in a total number of measures in excess of 40. A full list of Core Standards can be found in Appendix 3. Performance against each individual core standard was measured on a bi-modal scale of 0 for compliant and 1 for 'Not met'. The number of targets not met then determined overall performance on a four point categorical scale, as with the Existing National Targets, of Not met (worst) to Fully met (best).

New National Targets

New National Targets were Government measures of performance that reflected developing priorities as Trusts moved beyond the standards reflected in core and existing standards. A full list of New National Targets can be found in Appendix 3. Performance against New National Targets was as for the Existing

National Targets with each indicator being scored on a 3 point categorical scale of 3 for achieved, 2 for underachieved and 0 for failed. These scores were then summed and the total score, depending on the number of applicable targets determined the overall scoring on a four point scale of Weak (worst), fair, good and excellent (best).

The Quality of Services rating:

The final Quality of Services rating was derived by aggregating the scores against each of existing national targets, core standards and new national targets according to the matrix shown in Figure 7. In this figure the rankings have been changed into a four point numerical ranking as follows: Not met (ENTs and Core Standards) or Weak (New National Targets) =4; Partly met or Fair =2; Almost met or Good =3; and Fully Met or Excellent =4.

Figure 7: Derivation of the final Quality of Services rating

Measurement scale 1-4[35]	**Existing National Targets (12)**	**Core Standards** (24)	**New National Targets (NNTs)** (13)	**Quality of Services rating** **49**
Aggregated Performance score	1 OR	1		Weak
	2 AND	2		Fair
	3 AND	3 AND	3	Good
	4 AND	4 AND	4	Excellent

As can be seen the AHC required progressively greater coverage of activities in order to progress from a lower to a higher rating. This design feature reduced the incentive to concentrate resources on a narrow range of measures as was observed in the Star ratings where it was possible to obtain a high performance rating of two stars by concentrating resources on a narrow range of nine key Government targets. As can be seen from Figure 5 the equivalent rating of Good

[35] Performance across each dimension of Existing National Targets, Core Standards and New National Targets has been converted into a numerical score of 1 (worst) to 4 (best) as follows. Existing National Targets and Core Standards: 1=Not met; 2= Partly met 3=Almost met; 4 = Fully met. New National Targets: 1= Weak; 2= Fair; 3=Good; 4= Excellent.

requires not only good performance against the Existing National Targets but also against the Core Standards and the New National Targets.

4.4.3.2 Financial measures in the Annual Health Check: Use of Resources

In the Annual Health Check performance against financial measures was not aggregated with nonfinancial measures, as in the Star ratings, but was assessed separately. Further, in contrast with the Star ratings, which included only one financial measure, that of reported financial performance, the Annual Health Check included a range of measures in five main perspectives: financial reporting, financial management, financial standing (which incorporated the measure of reported financial performance), internal control and value for money.[36] Each individual measure was assessed on a four point scale of one to four by the Audit Commission (to whom this function was delegated by the Healthcare Commission) who provided the relevant information for determining a Trust's Use of Resources rating to the Healthcare Commission. These scores were then aggregated into an overall Use of Resources rating as in Figure 8.

Figure 8: Derivation of the Use of Resources rating

<table>
<tr><th></th><th>Financial standing</th><th>Financial Management</th><th>Value for Money</th><th>Financial Reporting</th><th>Internal Control</th><th>Use of Resources rating</th></tr>
<tr><td>Number of measures</td><td>2[37]</td><td>3</td><td>4</td><td>2</td><td>3</td><td>14</td></tr>
<tr><td rowspan="4">Performance Measurement scale 1-4</td><td>1 OR</td><td>1 OR</td><td>1</td><td></td><td></td><td>WEAK</td></tr>
<tr><td>2 AND</td><td>2 AND</td><td>2</td><td></td><td></td><td>FAIR</td></tr>
<tr><td>3 AND</td><td>3 AND</td><td>3 AND</td><td>2 AND</td><td>2</td><td>GOOD</td></tr>
<tr><td colspan="3">AT LEAST TWO SCORES OF 4 AND</td><td>3 AND</td><td>3</td><td>EXCELLENT</td></tr>
</table>

[36] The scoring methodology was published for each year of the Annual Health Check and can be obtained from: http://www.cqc.org.uk/guidanceforprofessionals/nhstrusts/annualassessments/annualhealthcheck2005/06-2008/09.cfm

[37] Financial breakeven is a critical indicator. Failure to breakeven results in a score of one for Financial Standing, thereby guaranteeing a WEAK Use of Resources rating.

.A list of the individual measures (14), most of which have subsidiary measures making up around 40 measures in total) can be found in Appendix 4.

Key features of this scoring system as compared with the Star ratings are the increased coverage of financial activities and the increased emphasis on the achievement of financial breakeven. As can be seen, similar to the Quality of Services rating, increasing coverage of activities combined with increasing levels of performance are required to progress from one rating level to another. Further, Financial standing in particular has a very influential place. There is only one measure for financial standing which is broken down into only two subsidiary measures, one of which is financial breakeven, the other being financial projections. Incurrence of a deficit of whatever size results in a financial standing score of 1 (the lowest) which then gives rise to an overall Use of Resources score of Weak, whatever the Trust's performance against the other measures. Further, in contrast with the Star ratings which only concerned itself with current year performance, in the Annual Health Check financial breakeven must be achieved in each of the previous three financial years if a score of 4 is to be achieved for financial standing. Both of these factors reduce the incentive to incur a short term tactical deficit in order to improve the quality of services rating and illustrate the continuing significance which is attached in the Annual Health Check to the financial breakeven objective.

4.4.3.3 Summary of key similarities between the Star ratings and the Annual Health Check

Although there are distinctive differences between the Star ratings and the Annual Health Check there remain some key common features.

1. Performance against a limited number of key Government performance measures (Key Targets in the Star ratings and the Existing National Targets in the Annual Health Check) was influential to the overall rating in both systems.

2. In each of the PMS, performance against each of these key Government measures was scored on a three point categorical scale relative to a predefined target.

3. The scores in all perspectives were aggregated using a scoring matrix to arrive at an aggregated performance rating measured on a four point scale of zero to three stars in the Star ratings and Weak to Excellent in the Annual Health Check.

4. Reported financial performance and the achievement of financial breakeven was an influential measure for the overall performance rating under both PMS

In addition, both PMS were linked to a number of performance management mechanisms aimed at incentivizing good performance ratings. These included rewards for good performance in the form of lower levels of monitoring, more commercial freedom including the ability to apply for Foundation Trust status and reputational gains (Department of Health, 2002c). Penalties for poor performance included the risk of dismissal and reputational damage.

In summary, the distinguishing features of the AHC, compared with the Star ratings, were a much wider range of measures, both financial and nonfinancial, and an increased emphasis on financial performance. However a number of common features characterise both systems and it is these common features which are exploited in the investigations in this thesis to derive uniform measures of financial and nonfinancial performance for all Trusts across the period of study.

4.4.4 Significance of the Star ratings and the AHC for research investigations

The Star ratings and the AHC were uniform systems of performance measurement that were applied to NHS acute hospital Trusts over the period 2003-2008. This setting, as in Ittner, Larcker and Randall's (2003) choice of the financial services industry, provides more homogeneous setting for exploring the relationship between nonfinancial and financial measures of performance (in this case, cost efficiency), than multi industry settings and mitigates the need to control for variations in incentives, organisation form etc. Further the availability of consistently

applied nonfinancial measures in the Star ratings and AHC mitigates the need for the use of self-reported measures of performance as, for example, in Hall (2008) and Hoque and James (2000). Further, the juxtaposition of the Star ratings and the AHC offers the opportunity for a quasi-experimental study, as in Propper, Sutton, Whitnall and Windjmeier (2010), which compares the impact of the AHC in comparison with the Star ratings on the relationship between nonfinancial performance and cost efficiency. Further, the two systems offer the opportunity of investigating the performance impact of increased measurement diversity and improved balance in the weightings given to financial and nonfinancial measures in balanced scorecard inspired PMS (Ittner and Larcker, 1998a) applied in a public sector setting, a setting which has received little attention in the empirical archival literature.

4.5 REPORTED FINANCIAL PERFORMANCE AND FINANCIAL SUPPORT: FLEXIBILITY AND MANIPULATION

4.5.1 Introduction

One common feature of both the Star ratings and the AHC is the importance attached to reported financial performance. In the Star ratings it represented one of only nine key Government targets that were influential in determining a Trust's ultimate performance rating, and in the Annual Health Check failure to achieve financial breakeven attracted the maximum penalty and the lowest rating for 'Use of Resources'. As has already been identified, reported financial performance and the break-even duty, represented key mechanisms for discharging Trusts' public accountability. However, during the 1990s underdeveloped costing systems, which impeded effective contracting, created an environment in which accounting mechanisms associated with the previous hierarchical regime and which offered flexibility in the reporting of financial breakeven, were retained. These included capital to revenue transfers which were withdrawn when the Government introduced accruals accounting to Government Departments; the ability to defer patient treatment, which became severely constrained when the Star ratings system was introduced and the system of financial support, in which surplus funds from

elsewhere in the NHS were transferred into NHS organisations in financial difficulty[38]. Other mechanisms such as accruals manipulation or the deferral of expenditure such as repairs and maintenance offered some scope for flexibility but were limited in their effectiveness for dealing with large deficits which might arise, for example, because of book losses on asset sales and as a consequence of heavy one off revenue expenditure such as might be associated with early retirement settlements or with large capital projects. In other instances, large deficits may arise as a consequence of poor management and the provision of financial support might be negotiated on the appointment to a Trust in financial difficulty of a new Chief Executive or Director of Finance.

This section of the thesis proceeds as follows. First the operation of the system of financial support is described. Second, the purpose of financial support is reviewed and the incentives not to apply it for the benefit of the public and patients but to use it opportunistically to avoid the consequences of performance measurement are considered. Third, the accounting policies relating to the provision of financial support are outlined. Finally the impact of the system of financial support on public accountability is reviewed. A number of case studies, taken from Audit Commission publications, are provided to illustrate certain aspects of financial support and these are referenced where appropriate.

4.5.2 How the financial support system worked

Financial support was an NHS specific transaction whereby surplus funds could be transferred as additional revenue into Trusts in financial difficulty. It was defined in the NHS accounting manual[39] as: 'additional income during the year, provided wholly to assist in managing financial problems' (Audit Commission, 2006a, para. 2.16). Financial difficulty is defined in the Department of Health's

[38] The longstanding availability of financial support was confirmed by Rt. Hon Kevin Barron MP, Chairman, Health Select Committee who, in the Health Committee's review of NHS Deficits (House of Commons, 2006b) referred to financial support (otherwise referred to as brokerage) as having been in existence 'for years and years' (p. Ev54).

[39] NHS Accounting Manual available at: http://www.info.doh.gov.uk/doh/finman.nsf/ManualDownload?OpenView

financial statements as being a deficit of more than £1m or more than 1% of revenue (as in the NHS Summarised Accounts, 2006 para. 42.[40])

Financial support could be sought from a number of sources. First, a Trust's main commissioner (PCT) might agree to provide additional funds from its own unallocated resources; second, a Primary Care Trust might seek to reallocate funds from other Trusts within their commissioning area, on the understanding that these funds would be returned in future years; third, funds might be sought from other Primary Care Trusts within the remit of the local Strategic Health Authority; and finally, funds could be sought from other (English) Strategic Health Authorities. From 2002, these largely informal arrangements were given a greater degree of structure when funding sourced from other Strategic Health Authorities became administered and co-ordinated by the newly created NHS Bank, essentially a mutual organisation of SHAs (Department of Health, 2003c) whose purpose 'is to support NHS organisations in maximising the use of resources across the NHS and over different financial years' (Audit Commission, 2005, p.49; 2006a, p.80). In this way the informal arrangements were institutionalised. The NHS Bank was allocated £100m of NHS central funds (the Special Mutual Assistance Fund) in each of the three years to 2005-06 (Department of Health, 2003c) to be allocated by way of a non-recurring grant to SHAs in need[41].

The NHS Bank also co-ordinated the brokerage of financial support beyond the £100m Special Mutual Assistance Fund between SHAs on a similar, but more formal basis than previously. Local arrangements for brokerage of funds between NHS organisations within an SHA continued, on an essentially informal basis, as previously. Where funds were needed to support NHS Trusts, the conditions upon which financial support was provided to the NHS Trust was subject to agreement between the Trust, the SHA and the local PCT. A general condition of the provision of financial support was the submission and agreement of a recovery plan which incorporated an agreed schedule of repayment of the funds advanced. Some

[40] NHS Summarised Accounts are available at: http://www.nao.org.uk/publications/0809/nhs_summarised_accounts_07-08.aspx

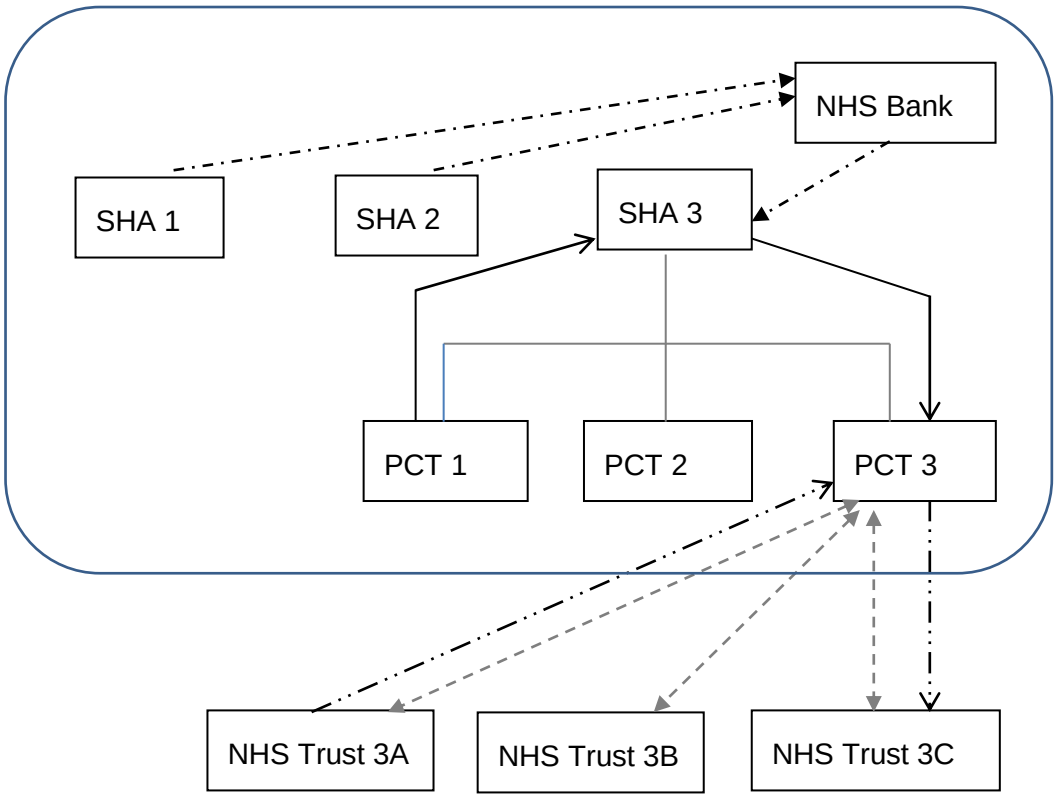
[41] The Mutual Support Assistance Fund essentially reflected a top slicing of the income allocated to PCTs and acted as a form of compulsory contribution towards the support of Trusts in financial difficulty.

categories of additional revenue costs associated with the early years of PFI projects were, however, routinely supported from NHS Bank funds on a temporary basis[42] through the provision of non-recurring grants.

The avenues through which funds could flow to NHS Trusts in the form of financial support are illustrated in Figure 9 below.

[42] Further information on the systems used to allocate financial support for PFI projects can be found in NHS Bank guidance for 2005-06.

Figure 9: Avenues for transferring funds into NHS Trusts as financial support



Accountability of PCTs 1, 2 and 3 to SHA 3

Service level agreements between NHS Trusts 3A, 3B and 3C with PCT 3

Boundary of the Department of Health

Transfers of funds, 'financial support' between SHAs, brokered by the NHS Bank, subsequently transferred to PCTs

Transfers between PCTs, brokered by the SHA

Transfers between NHS Trusts in the same commissioning area as PCT 3, brokered by the PCT

Financial support was available to both Primary Care Trusts and to NHS Trusts but whilst Primary Care Trusts operate within the departmental boundary of the Department of Health, NHS Trusts are quasi-independent organisations that operate outside the Department's accounting boundary. As such financial support provided to NHS Trusts was always provided via the local commissioning PCTs and not directly for example from other NHS Trusts. Recovery of financial support in future periods was also effected through the local PCT.

4.5.3 The purpose of financial support

Some indication of the purpose of financial support is given by its definition in the NHS accounting manual[43] as: 'additional income during the year, provided wholly to assist in managing financial problems' (Audit Commission, 2006a, para. 2.16). More bluntly the National Audit Office (2004, p19) describes it as 'a tool used by the Department (*of Health*) to help NHS organisations achieve financial balance'.

Further insight is given into its purpose by the proceedings of the Health Select Committee's investigation into the origins of NHS deficits (House of Commons 2006a, b). Phil Taylor, Chairman of the Healthcare Financial Management Association described the purpose of financial support as follows:

"If a trust was opening a new facility, in the first year or two it is much more expensive when you open a new facility and so you need to pass a little extra bit of funding to that organisation in order to get over that hump. There could be other reasons for moving brokerage *(financial support)* round the system, but the intention always was not to make the system less transparent but to oil the wheels to make the NHS able to cope with local difficulties".

The significance of deficits for patient care is illustrated in the following extract from the Audit Commission's review of the NHS Summarised Accounts for 2004-05:

"Managing and recovering deficits can have a far-reaching impact on NHS bodies, potentially affecting their ability to deliver services, meet binding

[43] NHS Accounting Manual available at:
http://www.info.doh.gov.uk/doh/finman.nsf/ManualDownload?OpenView

commitments, manage major initiatives and achieve the criteria for Foundation Trust status……. Pressure to recover a deficit and avoid breaching financial duties will mean NHS bodies are faced with difficult decisions, the results of which can impact on service delivery. This can include reducing capacity, for example through staff cuts, vacancy freezes or ward closures, or generating income through non-recurrent measures such as property disposals. Such measures, while providing temporary relief to financial pressures, may well impact on the body's performance against other key targets, such as access to services and waiting times" (Audit Commission 2006a p33).

The pressure to avoid a deficit was however strong. Ballantine, Forker and Greenwood (2008b) found that deficits were significantly associated with CEO turnover. Case study evidence is also provided by the Audit Commission's work on NHS financial statements. Case Study 1 describes the experience of the Mid Yorkshire Hospitals NHS Trust and illustrates how the Director of Finance was also under threat of dismissal. Further more desperate measures are illustrated by Case Study 2 which describes how the Scarborough and North East Yorkshire NHS Trust resorted to inappropriate accounting adjustments in an attempt to report financial break-even. Key points arising from the case studies are highlighted in bold and accompanied by explanatory text in an adjacent text box.

Case Study 1: The Mid-Yorkshire NHS Trust

In 2003-04, the Mid Yorkshire Hospitals NHS Trust reported an in-year deficit of £18.6 million, the largest deficit of all NHS bodies in 2003-04. The Trust and the local health economy have long-standing service and financial issues, which the Trust had been able to manage in previous years through **the receipt of one-off financial support** and a variety of other non-recurrent solutions. However, these measures did not address the underlying problems. The Trust, in conjunction with their auditors, has identified the main contributors to the deficit in 2003-04 as: failure to adequately manage recruitment and retention difficulties leading to the use of high-cost bank and agency staff; **the cost of meeting key NHS Plan waiting-time targets through premium-rate waiting list initiatives**, **including the use of the private sector**; and the historic lack of adequate budget-setting and monitoring procedures, making it more difficult to identify, monitor and manage cost pressures. Progress has been made in this area during 2003-04 and beyond, but by this time the deficit had already arisen.

Financial support to allow time to address financial problems

Trade off of financial for nonfinancial performanc

The timeline below shows the deterioration of the Trust's financial position

May 2003: Trust Board approves budget for 2003-04 which identifies savings requirement of £19 million to achieve financial balance.

June 2003: Year end deficit of £8.7 million predicted.

August 2003: **New Director of Finance takes up post.**

October 2003: Director of Finance's review of financial position shows that year-end deficit could be as high as £34 million.

January 2004: Trust agrees plan to reduce year end deficit to £18 million.

July 2004: The Trust's 2003-04 annual accounts show a deficit of £18.6 million (£30.6 million if external support is removed).

September 2004: Auditor issues Section 8 public interest report because in his opinion insufficient progress had been made on the actions identified in his February letter. The auditor stresses the urgency of agreeing a recovery plan with the Department and the Strategic Health Authority.

Consequences for finance director

Audit Commission 2005 p.16

Case Study 2: Scarborough and North East Yorkshire NHS Trust

Scarborough and North East Yorkshire Healthcare NHS Trust has had **financial difficulties for a number of years. In 2004-05, the Trust was in year three of a financial recovery plan agreed with its main commissioner and the local Strategic Health Authority. Over the last four years the Trust had received financial support totalling £10 million from the Strategic Health Authority; in each case, the Trust had repaid this support in the year following its receipt. With this support, the Trust met its key NHS Plan waiting time targets for each year and its financial break-even target until 2004-05.**

……

… during 2004-05 **the Trust also considered a number of further measures to break even. These included a number of accounting adjustments. Before the accounts were prepared, the appointed auditor provided guidance that the proposed adjustments would not comply with accounting standards as set out in the NHS Trust Manual for Accounts.**

The Director of Finance chose to disregard the auditor's view and prepared a balanced set of accounts. **The draft accounts submitted for audit in May 2005 reported a year-end surplus of £20,000 and contained a number of inappropriate accounting adjustments and errors**. Despite the existence of clear rules on large adjustments related to purchases made in previous years, the Trust hoped it could reduce in-year spending by reclassifying previously purchased medical instruments as stock and fixed assets. The accounts also contained examples of incorrect accounting treatment and inadequate checking procedures leading to significant errors. In the Public Interest Report, the auditor reported that **a number of adjustments employed by the Trust were a device to achieve financial balance, rather than improve the accuracy of the accounts, and did not comply with accounting standards.**

Audit Commission 2006a, p52

Financial support accompanied by recovery plan

Repayment of financial support

Financial support to allow time to address financial problems without damage to patient services.

Indication of the strength of the incentives to report financial breakeven

Finally Case Study 3 illustrates the concerns of the Audit Commission about the late provision of financial support in the Royal Cornwall NHS Trust.

Case Study 3: Royal Cornwall NHS Trust

Even where proper disclosure is made, unplanned **support is sometimes used as a last-minute 'fix' to prevent bodies breaching statutory financial duties**. In 2004-05, Royal Cornwall Hospitals NHS Trust received £15.5 million of unplanned support in addition to £10.9 million of planned support. This was precisely the figure required to eliminate its cumulative deficit of £13.6 million and hence meet its statutory duty to break even within four years (including an agreed extension). Although this arrangement was fully disclosed in the Trust's annual accounts, it is evident that the current regime allows bodies to receive financial support at short notice to avoid breaching statutory duties.

Audit Commission, 2006a, p.22

Opportunistic application of financial support

These case studies illustrate that the potentially opportunistic application of financial support required the collaboration of several layers of NHS management. The incentives for the avoidance of NHS Trust deficits lay in the different roles each organisation played. The SHA, in its role as performance manager for all NHS Trusts and PCTs in its area, may wish to avoid scrutiny of its possible failure to exercise this role effectively. Similarly, the PCT might wish to avoid questions about its own operations and allocation of funds to hospital services. If the PCT was struggling to balance its own books an adjustment to hospital service level agreements might serve only to transfer a deficit from the hospital to the PCT, and could set a precedent for higher future levels of funding, an outcome that might be particularly undesirable if there were concerns that hospital management was not pursuing realisable efficiency savings with sufficient vigour. Evidence of the opportunistic use of the financial support system was provided by Dr. Doug Naysmith MP as part of the proceedings of the Health Select Committee's review of NHS Deficits (House of Commons, 2006, b).

'My experience of that system *(financial support)* was that it was sometimes used to cover up deficits and then, at an appropriate time, the money was moved

back again to where it had come from without any real effort being put into sorting out why the deficit arose in the other organisation first.' (House of Commons, 2006b: Q476).

However, as is illustrated by Case Studies 1 to 3, the use of financial support without adequate recovery plans became an, increasing concern for the Audit Commission:

'Much has been written about such support and brokerage and the culture of dependence that this creates. We found that in failing organisations the dependence on brokerage was absolute, with recovery planning focusing to an unhealthy degree on the precise profiling of future borrowing and the schedule of repayments'. (Audit Commission, 2006b, para. 76).

Financial support was not however meant to allow flexibility in the reporting of breakeven where deficits arose in the context of routine trading transactions. Where deficits arose for example because of a higher level of patient referrals treatment than was anticipated in the Trust's service level agreement with the PCT, amendments to that agreement could be negotiated between the two parties. Case Study 4 illustrates a number of features of the system of financial support and the means by which Trusts could address deficits, including the renegotiation of service level agreements. Other features illustrated by the Case Study include: the provision of financial support to facilitate the reporting of financial breakeven; SHAs and PCTs as sources of financial support; the trading off of financial performance for nonfinancial performance prior to the receipt of financial support; the anticipated improvement in subsequent underlying financial performance; and the presence of a recovery plan and of process improvements which permit financial recovery whilst simultaneously improving service standards.

Case Study 4: North Bristol NHS Trust

In the year to 31 March 2003, North Bristol NHS Trust reported the largest ever deficit in the NHS of £44.6 million. **In the year to 31 March 2004, the Trust reported an in-year surplus of £20,000 after receiving £32.2 million of external financial support (£18.1 million from the NHS Bank and £14.1 million from Primary Care Trusts). Stripping out the external support, the Trust managed to reduce its deficit for the year from £44.6 million to £32.2 million, an improvement of £12.4 million**. The improvement was net of new cost pressures of £3.8 million, so the actual savings achieved were £16.2 million. The Trust notes that the main contributing factors to the reduction in deficit were: reduction in agency staff costs (£5 million); **reduction in expenditure on private sector treatment to meet access targets (£2 million);** reduction in depreciation charge (£1 million);**additional income received for exceeding service level agreement activity level (£1 million);**rates rebate (£1 million)…….. It plans to make further savings in 2004-05 of £19.3 million..… The largest individual contributors to the planned £19.3 million savings are: **£2.5 million from increasing the Trust's capacity in orthopaedics through efficiency improvements thereby reducing the use of the private sector and waiting list initiatives to treat patients…..£1.0 million from bed reductions through reducing the length of patients' stay in hospital**; £1.0 million from procurement savings……Features of how the Trust is managing its **recovery programme** include the following:…

- Realistic savings targets: Savings targets are set by taking account of the best available information on the scope for savings, including benchmarking information, reference cost comparisons, and known savings opportunities…

- **Improvements in patient care processes: There is a focus on sustainable improvements in patient care processes that increase quality of care as well as reducing costs**

Audit Commission 2005, p.27

Financial support allows Trust to meet breakeven requirement

SHAs and PCTs as source of financial support

Improvement in subsequent underlying financial position

Previous trade off of financial for nonfinancial performance

Renegotiation of service level agreement

Recovery facilitated by process improvements which simultaneously improve service standards

4.5.4 Accounting for financial support

Financial support was credited in the books of the receiving Trust as revenue (Audit Commission, 2005). As such it was treated in the books of account in a way which was consistent with its definition as 'additional income during the year'. However, consistent with the tension between hierarchical and market modes of governance, and the origins of financial support in the budget reallocations between cost centres within a hierarchy represented by the local health authority, the accounting transaction of 'financial support' between PCTs and NHS Trusts was characterised by a high degree of ambiguity.

The majority of financial support transactions took place on the basis that a surplus of one organisation would temporarily be released for the benefit of another organisation in the expectation that the funds transferred would be 'repaid' in future years. This, for example, is illustrated in Case Study 2: Scarborough and North East Yorkshire NHS Trust. Prima facie, therefore, these transactions had the characteristics of 'loans'. However, unlike a loan, financial support was credited to revenue rather than liabilities. Further in a throwback to hierarchical modes of governance where hospitals were treated as cost centres, repayments were effected not by an expense line in the income and expenditure account, but by the top slicing of revenue agreed under the terms of the following years' service level agreements[44]. From a hierarchical perspective this treatment can be regarded as being similar to the recovery of previous budget reallocations. However, from a market based perspective financial support can also be seen to result in an advancement of revenue from a future period to the present and to be similar in its effects on reported financial performance as the manipulation of revenue accruals in the private sector. An illustration of this effect can be found in Figure 10 below.

[44] See for example, the NAO's report on the NHS Summarised Accounts for 2002-03 which describes the movement of financial support between NHS organisations as follows: "Underspends in other parts of the health system may be used to provide additional resources to those NHS organisations that require additional funds. This is via an adjustment to a service level agreement and the increase in funding may be reversed in the following year subject to the discretion of the Strategic Health Authority. (National Audit Office, 2004, p20)

Figure 10: Case study illustration of the impact of financial support on reported revenue

A hospital Trust has revenue of £200m per annum and has historically broken even in each and every year. It is about to open a new accident and emergency unit and the additional one-off costs of transition from the old to the new unit are estimated at £6m. The Trust therefore has a choice between maintaining service standards and incurring a deficit or breaking even and allowing service standards to deteriorate, for example by extending waiting times for treatment. To avoid the adverse consequences to patients of reduced service standards, or the adverse reputational damage of a deficit, management could, as an alternative, seek financial support of £6m to allow the Trust to breakeven and maintain service standards. The Trust management prepares a case for financial support and agrees to repay the £6m over a period of two years out of the efficiency savings which are anticipated from the new unit.

Thus the revenue of the Trust appears as follows:

	Year 0	Year 1	Year 2	Year 3	Year 4
Earned revenue	£200m	£200m	£200m	£200m	£200m
Financial support	£0m	£6m	-£3m	-£3m	£0m
Reported revenue	£200m	£206m	£197m	£197m	£200m

4.5.5 Impact on accountability

The effect of financial support on contemporaneous reported financial performance was unambiguous: it served to reduce or eliminate an underlying deficit allowing Trusts in many cases to report financial breakeven in their financial statements and to fulfil their statutory duty to breakeven. The requirement to produce audited, true and fair financial statements has represented a key element in the public accountability of NHS Trusts since their inception and for these purposes the treatment of financial support as revenue was, in the context of the NHS setting, interpreted as being UK GAAP compliant (HM Treasury, Financial Reporting Manual, Chapter 2; NHS Accounting Manual). As a consequence, financial support often resulted in improved performance ratings and the avoidance of a breach of the breakeven duty. The accounting treatment of financial support meant that there was an increase in reported revenue in the year(s) that it was received and a reduction

in reported revenue in years that it was ‘repaid’. The reverse occurred for Trusts releasing funds for brokerage as financial support to Trusts in financial difficulty. They reported reduced revenue in the year of release and increased revenue in the year(s) of ‘repayment’.

Prior to 2004 there was no requirement to disclose financial support in Trust financial statements with the result that its impact was completely hidden. However, following the Audit Commission’s repeated concerns about the impact of financial support on Trust accountability (Audit Commission, 2005) those Trusts in receipt of financial support were required, from 2002-03, to disclose the amount received and whether it was sourced from the NHS Bank or from the local health economy. Compliance with this requirement was however patchy in 2002-03 and it was not until 2003-04 that the standard of reporting improved. Notwithstanding its material impact on Trust financial performance, the disclosure of financial support was by way of note (Note 7, NHS Accounting Manual, Chapter 7) rather than on the face of the income statement. Further, the reporting of financial support was asymmetrical in that ‘repayments’ were not subject to a disclosure requirement nor were Trusts which released funds for brokerage required to make equivalent disclosures.

Disclosure did not, however, affect the flattering effect of financial support on the satisfaction of the statutory duty to breakeven and it continued to reduce the number of breaches of the duty reported to Parliament. Similarly Trust performance ratings, which were held out as a measure by which the public could judge the performance of a hospital Trust, also continued to be flattered.

However, where such transfers offered little threat to the Department of Health’s primary accountability of overall breakeven across the healthcare system there were potential benefits to patient care in smoothing income and expenditure across individual organisations and across time as a means of managing the recovery of Trusts in financial difficulty. The institutionalisation of financial support through the approved accounting policy, which was considered appropriate to the NHS setting, and through the setting up of the NHS Bank to co-ordinate major transfers across regional health economies is an indication of a wider institutional perception of such benefits. However, in 2006 the Department of Health acknowledged that the accounting treatment of financial support was serving to

mask underlying financial problems (Audit Commission, 2006a; House of Commons, 2006a, Department of Health,2007).

In 2006-07 the system of financial support was abandoned in favour of leaving deficits where they arose. This move however was not primarily driven by concerns about accountability, which was problematic even after the disclosure requirements introduced in 2002, but by an increasing momentum towards market based mechanisms. These included the introduction of Foundation Trusts, for which there was no access to financial support, and the opening up of hospital service provision to private sector providers. The cash problems that potentially arise from a Trust being in deficit are now addressed by a system of interest bearing loans[45] (Audit Commission, 2006a), mimicking the system of resource access that is available to Foundation Trusts and commercial providers. As these funds are not credited to revenue but to liabilities, the accountability issues associated with the system of financial support are addressed. However, from an incentive perspective, by not being credited to revenue the new system does not address the possible adverse consequences of Trusts attempting to report financial breakeven through short term cost cutting. From a wider system perspective, the achievement of overall financial balance across a health economy, a key performance indicator for SHAs, is now facilitated through a contingency reserve held by the SHA. This reserve is held centrally by the SHA and not allocated out to individual Trusts whose deficits are reported where they arise (Department of Health, 2007). Thus the potential benefits to patient treatment in NHS Trusts arising from the flexibility afforded by financial support in achieving financial balance are no longer available.

4.5.6 Financial support: summary

In the NHS, the system of financial support operated to transfer funds into Trusts in financial difficulty from surplus funds elsewhere in the NHS. This system, which permitted flexibility in the reporting of financial break-even for Trusts in

[45] Such a system, formerly referred to as cash brokerage, has always run alongside the system of financial support, to address liquidity problems. Transfers of cash into a Trust receiving cash brokerage were debited to cash and credited to liabilities. This is in contrast to financial support which was otherwise known as ‘resource brokerage’ and was credited to revenue rather than liabilities. The current system of interest bearing loans replaces the previously informal arrangements to transfer cash between NHS organisations.

financial difficulty, was deeply embedded in the institutional structure and history of NHS and its evolution from hierarchical towards market modes of governance. As a form of income shifting it acted both across the NHS and, for individual organisations, across time by advancing the recognition of revenue. Manipulation of revenue has been identified as a common form of earnings management in the private sector (Dechow and Schrand, 2004) and has also been identified in the public sector (Vinnari and Nasi, 2008). Consistent with Demski, Frimor and Sappington (2004), financial support was applied with the intention of delivering benefits to patients by avoiding the adverse consequences of addressing deficits through short term cost cutting to achieve financial breakeven. The system of financial support was applied selectively and was not available to all Trusts in deficit. Consistent with Demski (1998) and Demski, Frimor and Sappington (2004), where benefits arising from income manipulation only arise where flexibility in reporting is restricted to 'good' managers, the requirement to produce a recovery plan served to act as a selection mechanism such that only highly skilled managers with credible recovery plans were given access to financial support.

Financial support was a form of income shifting that did not affect the overall consolidated results of the NHS but did affect the reported financial performance of NHS Trusts, for whom reported financial performance and Trust performance ratings were key mechanisms for discharging their public accountability. In this respect the setting is similar to that found for Japanese quoted companies in Thomas, Herrmann and Inoue, (2004) where parent company financial statements were manipulated, consistent with internal income shifting, even though there was no impact on the corporate financial statements.

The accounting treatment for financial support, although compliant with the official guidance in the NHS Accounting Manual, served to disguise the impact of financial support on the underlying financial performance of NHS Trusts and to impair their accountability in a way which was similar to that found in Finnish municipal authorities (Vinarri and Nasi, 2008). The complex interaction of incentives and accountability within the NHS, however, also led to concerns that financial support was being applied opportunistically and not generating benefits in the form of higher standards of service. Rather, it was being used to flatter reported financial performance and Trust ratings and to avoid breach of the statutory duty to breakeven. Although there is some anecdotal evidence of this there is no systematic

empirical evidence at present to identify whether such opportunistic application of financial support was significant or whether, on balance, it did deliver benefits to stakeholders in the form of higher service standards.

Despite the formal institutionalisation of the system of financial support, as evidenced by the role of the NHS Bank, the flexibility in the reporting of financial breakeven was withdrawn in 2006-07, not because of the concerns around its impact on accountability, which was problematic throughout its existence, but because of the need to demonstrate that all hospitals, whether public or private sector, were being remunerated for their services in an equitable manner. (Audit Commission 2006a)This issue became particularly pertinent on the opening up of the market for hospital services to private sector providers in 2006.

4.6 SUMMARY

This Chapter sets out the institutional context which provides the setting for two investigations of this thesis. It reviews the position of NHS Trusts within the wider setting of the commissioning and delivery of hospital services in England, and summarises their accountability to Parliament and the public for their financial performance, particularly the fulfilment of the statutory duty to break even, through the preparation of true and fair accruals based financial statements. It also sets out the context in which nonfinancial performance and cost efficiency, both of which are matters of public concern, are incentivised and controlled through mechanisms, such as multi-dimensional performance measurement systems and cost efficiency measures, which, particularly in the case of the performance measurement systems, provide a basis for performance management.

The characteristics and features of each of the Star ratings and the Annual Health Check are outlined and their differences and similarities highlighted. Specifically the wider measurement diversity and greater balance between financial and nonfinancial measures in the Annual Health Check as compared with the Star ratings is identified. These characteristics facilitate an investigation of a key policy aim of these systems, the relative effectiveness of Annual Health Check as compared with the Star ratings in incentivising cost efficiency as a means of delivering quality hospital services. This question is of interest to policy makers, service users and taxpayers. The juxtaposition of these two PMS also offers the

opportunity of investigating the relative impact of PMS design on the opportunities for manipulation. In this thesis the focus is on the manipulation of financial measures and, in the first study, the differing incentives built into each PMS facilitate an examination of their impact on the propensity to trade financial for nonfinancial performance.

Reported financial performance represented the basis for determining whether a Trust had fulfilled its statutory duty to breakeven, and was also influential in determining a Trust's performance rating in both the Star ratings and the AHC. In the final section of this Chapter, a system of transactions, known as financial support, which shifted income between NHS organisations and which was deeply embedded in the institutional structure and history of the NHS, is reviewed. Financial support allowed a degree of flexibility in reported financial performance and the fulfilment of the statutory duty to breakeven for Trusts in financial difficulty by reducing or, in most cases, eliminating deficits. Its purpose was to avoid the adverse consequences to patient services of short run tactics to report financial breakeven but the complex multi-layering and evolving interactions of incentives within the NHS led to concerns that it was being used opportunistically to disguise poor financial performance without generating any associated benefits. In the second study in this thesis the impact of the system of financial support on service standards and on Trust accountability is investigated.

CHAPTER 5 RESEARCH DESIGN

5.1 INTRODUCTION

This chapter comprises three main sections. Section 1 reviews the research methodology adopted for this thesis and the following two sections proceed to outline the research design for each study.

5.2 RESEARCH METHODOLOGY

This thesis adopts a positivist paradigm informed by an agency perspective. The Government has, since the early 1990s, been progressively moving hospital management from hierarchical modes of governance to market based modes of governance. This process has brought agency relationships increasingly into focus. It is appropriate therefore that an approach which is consistent with the policy direction of the Government is used to evaluate its policy initiatives. Further, although infrequently used in the accounting literature, this approach is common in the health economics literature which has a strong interest in performance measurement and management. Studies adopting this perspective to investigate the impact of performance targets on service delivery include, for example, Propper, Sutton, Whitnall and Windjmeier, (2010) and Kelman and Friedman, (2009) both of which generate findings of interest and relevance to the interpretation of the two studies in this thesis. The adoption of this research paradigm also permits the exploitation of a wealth of publicly available data, generated by the Government's policy initiatives which is to date, relatively unexplored (Goddard, 2010).

The adoption of an interpretive research paradigm is an equally valid approach to the investigation of performance measurement. This approach generates different insights into the impact of performance measurement systems, about how they are used, and about the institutional and political context in which they are developed and implemented. Institutional theory, for example, has been used as a vehicle to interpret the impact of performance measurement in the NHS by, for example, Ballantine, Brignall and Modell, (1998) and by Chang (2006, 2007, 2009). Ouchi's (1979, 1980) modes of governance (markets, hierarchies and clans, were also used by Aidemark (2001) to investigate the application of a balanced scorecard in Swedish hospitals.

Interpretive studies generally adopt interview, survey and case study methods as vehicles for their investigations. However, such methods would present considerable challenges for the purpose of investigating organisation wide performance characteristics such as Trust cost efficiency, an assessment of which requires systematic collection of data and an appropriate measurement metric. Notwithstanding this, there is scope for investigating whether perceptions of cost efficiency are aligned with the data. The Department of Health's investigations (2010) suggest, for example, that perception of data quality for Trust cost efficiency is based on early experience of the system of reference costs rather than reflecting the current state. Perceptions of levels of hospital efficiency might similarly be at odds with the evidence of Reference Cost Indices.

Although not explored in this thesis, alternative research paradigms are potentially useful to investigate the evolution of policy towards gaming and manipulation of performance measurement in the English NHS. The application of qualitative paradigms would be appropriate to further investigate, for example, official policy responses to flexibility in reporting financial performance and the evolution of accounting policy in response to NPM initiatives.

5.3 STUDY 1: PERFORMANCE MEASUREMENT SYSTEM DESIGN AND THE INCENTIVISATION OF COST EFFICIENCY

The first study of this paper investigates the relationship between cost efficiency and service standards during the period of study and the relative effectiveness of the AHC as compared with the Star ratings in incentivising cost efficiency as a means of improving service standards.

5.3.1 Introduction

A key concern of policy makers, taxpayers and users is the efficiency with which health services are delivered. Efficiency is important because for a given level of resources a higher level of efficiency permits more and better services to be delivered. Increasing life expectancy and developments in medical technology have for many years created increasing demands for health services and a corresponding demand for increased efficiency to offset the potential increases in costs. However a focus primarily on financial performance during the 1980s and 1990s led to

concerns about service quality, particularly access to care. The Labour Government elected in 1997 remained committed to the delivery of cost effective services, introducing and developing reference costs and Trust reference cost indices from 1998, but also made a commitment, in the NHS Plan, to improve service quality (Department of Health, 2000, p.5). A key mechanism for incentivising the cost efficient delivery of healthcare was the Performance Assessment Framework (Department of Health, 1999, p.1). The Star ratings system was introduced in 2001-02 and was replaced by the Annual Health Check in 2005-06. This juxtaposition of these two performance measurement systems provides the setting for an investigation of the relationship between service standards and cost efficiency and, further, which of these two systems was most effective at incentivising cost efficiency. This is the subject of this first study of this thesis.

5.3.2 Research design

The research design for this study draws on both the NHS Plan (Department of Health, 2000) and the Performance Assessment Framework (Department of Health, 1999), both of which make a commitment to drive up service quality through improved efficiency. A key feature of this study is the characterisation of the objective function of NHS Trusts as the maximisation of nonfinancial performance in the context of a financial resource constraint represented by its revenue, negotiated with its commissioners in service level agreements. Thus nonfinancial performance is characterised as being a function of cost efficiency, subject to the achievement of financial breakeven.

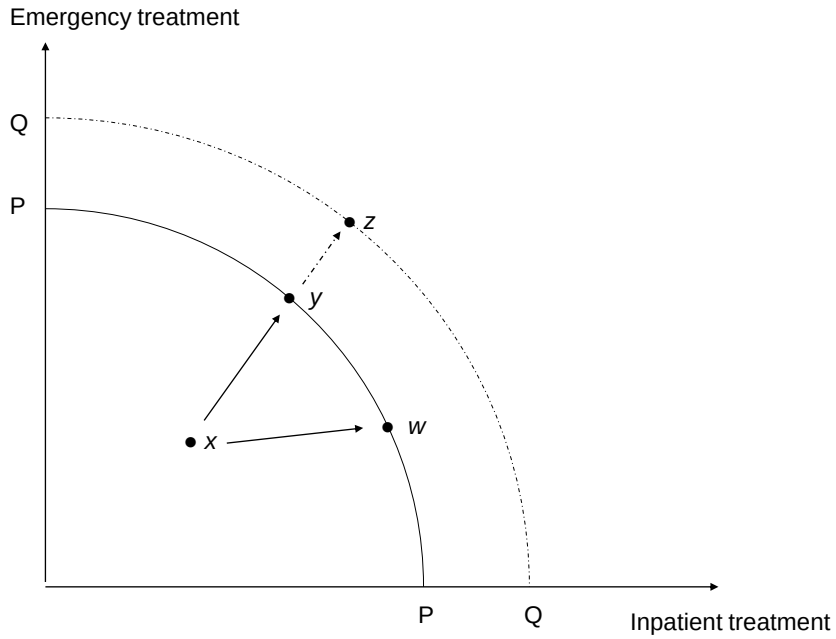
$$NFP = f(e, r) \qquad (1)$$

Where: NFP is a measure of nonfinancial performance; e is a measure of Trust cost efficiency and r is a measure of reported financial performance, which measures the extent to which there is a variation from financial break-even.

This characterisation of the objective function permits the use of production possibility frontier analysis to derive predictions about the extent to which the AHC as compared with the Star ratings incentivised cost efficiency as a means of improving service standards. These predictions are derived by considering the output decisions of a stylised hospital Trust with a limited number of outputs. First,

the output decision of a hospital delivering a combination of two outputs, emergency treatments and elective inpatient treatments is considered by reference to Figure 11.

Figure 11 : PMS design – incentivising hospital cost efficiency (1)



PP is the production possibility frontier of efficient combinations of two outputs in the presence of a financial resource constraint. *y* is the efficient goal congruent combination of outputs for a Trust breaking even and maximizing stakeholder utility; *x* is an inefficient combination of outputs for a Trust breaking even; *z* is the combination of outputs that would be delivered by an efficient Trust if the resource constraint were relaxed to QQ. *w* is a combination of outputs on PP which is efficient but not goal congruent. The incentive effects of PMS design on an inefficient Trust at x can be summarized:

A. A Trust subject to a PMS based only on an aggregated measure of reported financial performance would remain at *x* and not be incentivized to move to *y*.

B. A PMS based only on an aggregate measure of nonfinancial would incentivize movement towards the combination of outputs *y*, either by increasing efficiency or by relaxing the resource constraint, and temporarily going into deficit, to a point where an efficient Trust would deliver a combination of outputs *z*.

C. A PMS based only on an aggregate measure of cost efficiency would incentivize movement towards PP but not necessarily to *y*. An output combination at *w* for example would maximize the Trust's performance rating but not be goal congruent.

In Figure 11 efficient combinations of these two outputs are represented by the production possibility frontier PP. However, in the absence of a stakeholder utility function and the existence of multiple stakeholders, there is little to guide the Trust as to the combination of outputs that maximises stakeholder utility. In such a setting a PMS such as the Star ratings or the AHC, which were both subject to extensive public and professional consultation, can serve to articulate stakeholder objectives. Trust performance ratings can then be considered as a proxy for the extent to which stakeholder objectives are met and, when linked to performance management mechanisms, act to incentivise goal congruent outcomes. Thus the Trust producing the efficient combination of outputs *w* , in Figure 11, would receive a lower rating than a Trust with the goal congruent output combination of *y* and the Trust CEO would be incentivised to transfer resources to emergency services from elective inpatient services. Points within the frontier, such as output combinations *x*, are possible but not efficient. In the absence of a multidimensional performance measurement system a Trust with this combination of outputs is meeting its financial objective of breaking even but is not incentivised to improve its efficiency and increase its output to combination *y*. Activity levels and service standards will remain inferior. This was essentially the situation which in the 1980s and 1990s when regulation was largely restricted to financial performance. As a consequence service standards, particularly access to care, remained a cause for public concern and when the Labour Government came to power in 1997 the reduction of waiting lists was identified as being the public's top priority for the NHS (NHS Plan, 2000). If however a performance measurement system is introduced into this scenario which, in addition to the achievement of financial balance, specifies the desired combination of outputs *y* and links this combination to performance incentives then a Trust with output *x* will be incentivised to move to *y* by improving its efficiency. Balancing the incentives in the performance measurement system is however important. If, in the overall performance rating, the achievement of financial balance is given a lower weighting than the production of output combination *y,* then a short term perverse incentive is created whereby the Trust's performance rating can be improved by sacrificing financial performance in order to improve nonfinancial performance. Thus, a Trust producing *x* could move to *y* not only by improving cost efficiency but also by dysfunctionally moving into deficit, effectively relaxing the resource constraint to position QQ where an efficient Trust would produce output equal to *z*. In the NHS, this strategy can only provide short term benefits as the

incurrence of persistent deficits attracts considerable reputational damage. The heavy emphasis on nonfinancial measures in the Star ratings suggests that performance ratings could be improved by the incurrence of a deficit to improve performance against nonfinancial measures (Bevan, 2006). Figure 12 illustrates a number of strategies whereby a Trust could achieve such an outcome.

Figure 12: Strategies for enhancing a Trust's Star rating by incurring or increasing a deficit

Total penalties on key government measures[#] (excluding reported financial performance)	**Maximum achievable Trust Star rating given reported financial performance of:**		
	Breakeven or better (0 penalty points)	Small deficit <=1% total revenue[46] (2 penalty points)	Large deficit >1% total income (6 penalty points)
0	3*	3*	2*
2	3*	2*	1*
4	2*	2*	1*
6	2*	1*	1*
8	1*	1*	0*
10	1*	1*	0*
12	1*	0*	0*

A
B
C
D

[#] Two penalty points were incurred for underachievement and six for significant underachievement against each performance standard.

Strategies to improve performance rating:

A. Incur expenditure to ensure achievement of non-financial targets, such as waiting times, and incur a small deficit (of up to c. £2m)

B. For Trusts already in deficit, incur expenditure to ensure achievement of non-financial targets, such as waiting times, and increase deficit up to 1% of revenue (c. £2m)

C. For Trusts already with a small deficit, incur expenditure to ensure achievement of nonfinancial targets, such as waiting times, and increase deficit to more than 1% of revenue (c. £2m)

D. For Trusts already with a large deficit of more than 1% of revenue, incur further expenditure to ensure achievement of non-financial targets, such as waiting times, and increase deficit without limit.

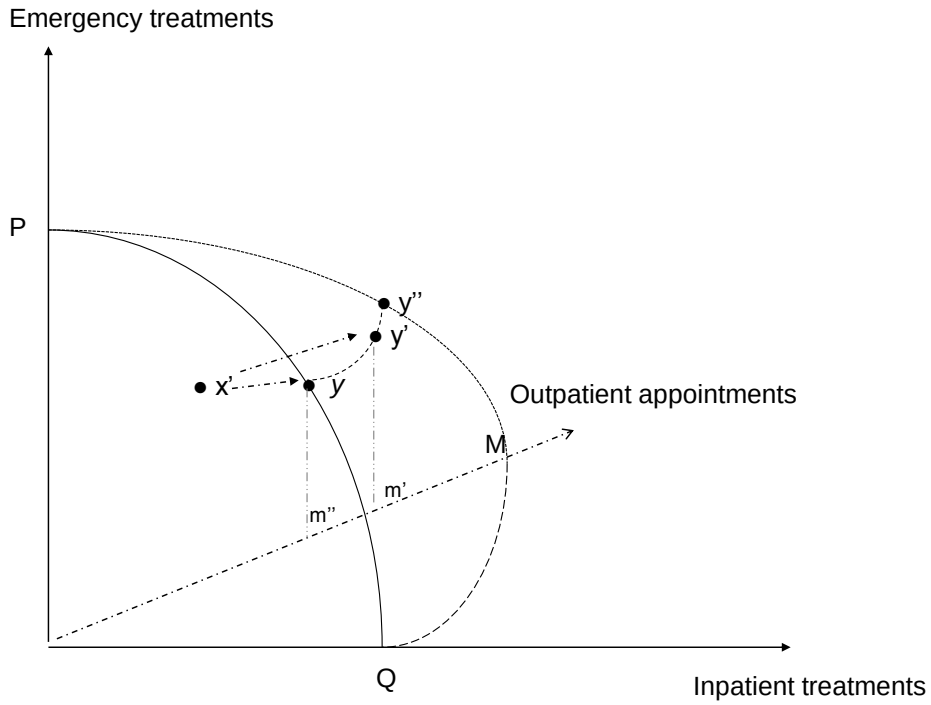
[46] 1% of revenue represents approximately £2m for a Trust with average revenue.

Taking Strategy A in Figure 12 as an illustration, a Trust which is expecting to breakeven but to be awarded only one star because it is expecting not to meet one or more waiting targets, could invest in waiting list initiatives, such as putting on extra theatre sessions. Even if the cost of these initiatives takes the Trust into deficit, its overall performance rating can be improved, to as much as three stars. Similarly, if a Trust is already in deficit, strategies B, C and D, which serve to increase the deficit, also result in an improved rating and once a deficit exceeds the threshold of 1% of revenue (i.e. about £2m for a Trust with average revenue) no further penalty is incurred. An additional incentive for the use of this strategy, was the potential availability of financial support, which, if approved, could serve to eliminate the adverse consequences associated with a deficit. In the AHC, in comparison, the incentive to incur a deficit to improve the performance rating was significantly reduced by isolating financial measures from nonfinancial measures, giving the lowest 'Use of Resources' rating for a deficit however small, and by requiring persistence in financial breakeven over a period of three years in order to achieve the highest rating. Further, under the AHC financial support was only available in its first year of operation, thus further reducing the incentive to overspend.

On the basis of this stylised example it is predicted that the AHC, will incentivise cost efficiency more effectively than the Stars by reducing the power of the incentive to go into deficit to achieve nonfinancial performance standards.

To progress this analysis further an additional output, outpatient appointments, is added to the stylised production possibility frontier. In the presence of this third output the production possibility frontier becomes the plane identified by PQM (Figure 13).

Figure 13: PMS Design – incentivising cost efficiency (2)



PQM is the production possibility frontier of efficient combinations of three outputs in the presence of a financial resource constraint; *y'* is the goal congruent efficient combination of these three outputs, associated with outpatient output m', for a Trust breaking even; *y* is a non-goal congruent efficient combination of the three outputs, associated with outpatient output at m''; *x'* is an inefficient combination of the three outputs. A PMS which includes all three measures of output will incentivize movement from *x'* to *y'*. A PMS which includes only inpatient and outpatient measures, will the highest rating to the combination of outputs which lies on the line yy''. For inefficient Trusts a high performance rating may therefore be achieved by moving through improved cost efficiency to y' or less efficiently by, effectively, reducing the output of outpatient appointments to m''.

In Figure 13, a PMS which fully reflects stakeholder priorities by incorporating measures for all three outputs, awards the highest performance rating to output combination *y'* where outpatient appointments are equal to m'. An inefficient Trust with output combination *x',* which operates within its financial resource constraint, will be incentivised by this PMS to improve cost efficiency and move to *y'*. However, if the PMS is only partial in its coverage of outputs, measuring only inpatient and emergency outputs, then any point on the line *yy''* will attract the highest performance rating, including output combination y where the output of outpatient appointments, at m'' is lower than m'. This analysis reflects the adage, attributed to Peter Drucker, that what gets measured gets managed and that good performance on measured activities may be at the expense of those not measured. For the Trust at x' the efficiency improvements required to maximise the performance rating by moving to y are less than the efficiency improvements required to move to y' because, effectively, resources have been transferred out of outpatient appointments (reduction in output from m' to m'') and into elective and emergency treatments. The Star ratings were subject to considerable criticism (Bevan and Hood, 2006) regarding the extent to which a Trust's performance rating could be improved by concentrating on a narrow range of activities represented by the Government's key targets as compared with the much broader based balanced scorecard indicators. However in the AHC the increased coverage of activities required to achieve the highest performance ratings reinforces the previous prediction that the AHC will incentivise cost efficiency as a means of improving service standards more effectively than the Star ratings.

In summary, the two distinctive characteristics of the AHC: increased measurement diversity and a better balance between financial and nonfinancial measures, signal a reduced potential for manipulation. First, the potential for improving a Trust performance rating through the trading off of financial for nonfinancial performance is reduced. Secondly, the potential for improving ratings through the concentration of resources on a narrow range of activities to the detriment of those not measured or with a lower weighting is also reduced.

Prediction: The AHC will incentivise cost efficiency to achieve performance standards more effectively than the Star ratings.

5.3.3 The performance comet

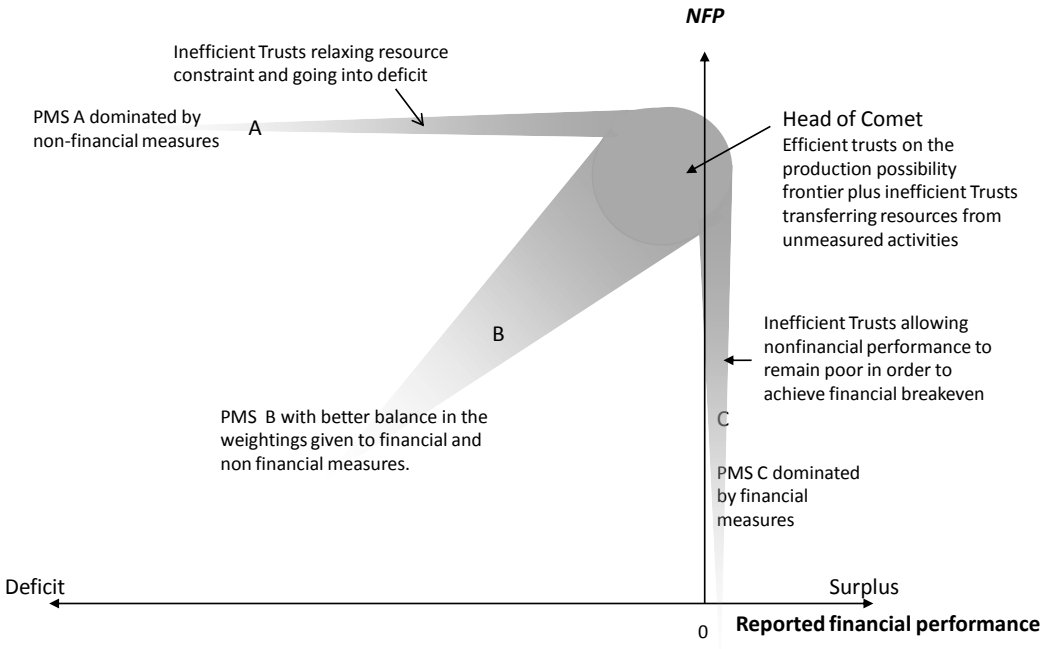
The two key features of the AHC as compared with the Star ratings that have been identified through previous analysis are, first, an increased emphasis on financial measures which result in a better balance between financial and nonfinancial measures, and secondly, increased coverage of activities which reduce the potential for neglecting unmeasured activities. As an aid to visualising the impact of these differences on the observed distribution of performance against nonfinancial measures and reported financial performance, a stylised illustration is provided in Figure 14. Figure 14 predicts that, in the presence of a PMS dominated by nonfinancial measures, (PMS A), Trusts operating efficiently on the production frontier and those which transfer resources from unmeasured activities will form a cluster of Trusts recording high levels of nonfinancial and financial breakeven. These Trusts are positioned in the 'head' of the comet. For PMS A there will also be a tail of Trusts recording high levels of nonfinancial performance but reporting a deficit. This tail comprises inefficient Trusts that maximise their performance rating by incurring a deficit in order to improve performance against nonfinancial measures. The Star rating system, by weighting performance against limited number of nonfinancial measures, much more highly than reported financial performance can be considered to be analogous to PMS A. PMS C on the other hand, is dominated by financial measures and therefore the 'tail' is characterised by Trusts which report financial breakeven but much more variable nonfinancial performance. This tail would comprise inefficient Trusts which achieve breakeven by allowing service standards to deteriorate. The performance measurement regime of the Thatcher era in the UK could be regarded as being similar to PMS C. PMS B has a better balance between the weightings given to nonfinancial measures and reported financial performance and also has greater coverage of activities. As the potential for resource transfer from unmeasured activities is reduced, the head of the comet comprises only efficient Trusts operating on the production frontier. All inefficient Trusts are now to be found in the tail of the comet. By virtue of the improved balance between financial and nonfinancial measures, the benefits of trading off financial more nonfinancial performance is more difficult to compute and the distribution of Trusts within the tail is therefore more dispersed. The AHC, as compared with the Star ratings is characterised by additional measurement diversity and a better balance between the weightings given to reported financial performance and nonfinancial performance. As a consequence, in comparison with

the Star ratings, the performance comet predicts that the distribution of nonfinancial performance and reported financial performance in the AHC will be more like PMS C than PMS A.

Multi-dimensional performance measurement systems have been applied to hospital systems not only in England but also in a number of other countries, notably Denmark (Aidemark, 2001) and Canada, an early innovator in balanced scorecard inspired measurement systems. The performance comet analysis is extended in Figure 15 which illustrates, in an elemental fashion, the positioning across the spectrum from A to C of a number of performance measurement systems that have been applied internationally to hospital systems.

Figure 14: The Performance Comet

An illustration of the impact of PMS design on observed performance against nonfinancial standards and reported financial performance in a public sector setting.



NFP is a measure of nonfinancial performance.

The Performance Comet illustrates the predicted distribution of reported financial performance and performance against an aggregate measure of nonfinancial performance in the presence of three PMS: PMS A is dominated by nonfinancial measures and offers scope for improving the performance rating by incurring a deficit and from the transfer of resources from unmeasured activities; PMS B is characterised by better balance in the weightings given to financial and nonfinancial measures and reduced scope for transferring resources from unmeasured activities; PMS C is dominated by financial measures and incentivises Trusts to report financial breakeven whilst allowing nonfinancial performance to remain variable.

Figure 15: PMS design - International comparisons

Illustrative examples of the location of international multidimensional healthcare PMS on the Performance Comet 'spectrum'

	SIMILAR TO PMS IN OPERATION IN:	**TIME PERIOD**	**CHARACTERISTICS:**
PMS unbalanced in favour of nonfinancial performance	World Health Organization: Performance Assessment Tool for Quality Improvement in Hospitals (PATH)	From 2003	Pilot performance framework with primary aim to facilitate international comparisons of hospital care quality and intended to sit within national frameworks. (World Health Organization Europe, 2003a,b; Arah et al., 2003; Groene et al., 2008)
	English Star Ratings System	2002-2005	Comprised c.40 measures, all nonfinancial with the exception of one financial measure, financial break-even. Dominated by nine key government measures. Performance aggregated into a single score ranging from zero to three stars
	English 'Annual Health Check'	From 2005	Two separate aggregated categories of performance: 'Use of Resources' (5 financial performance measures) and 'Quality of Services' (c. 70 measures).
	Canada	from 1999	Ontario Hospital Association: Hospital Report, a BSC inspired PMS covering 146 hospitals and comprising more than 30 indicators over four perspectives: system integration and change (12 indicators), patient satisfaction (4 indicators), clinical utilization and outcomes (9 indicators) and financial performance and condition (9 indicators). (Figures in parentheses relate to 2007). (Canadian Institute for Health Information, 2007; Pink et al, 2001; Groene et al., 2008; Arah et al., 2003).
PMS unbalanced in favour of reported financial performance	UK Patient's Charter	1990s	Emphasis on financial measures. Patient's Charter measures were nonfinancial but were largely outside control framework. (Mayston, 1985).
	UK	1980s	Heavy emphasis on financial measures.

5.3.4 Data and sampling

To investigate the relationship between service standards and efficiency and whether the AHC better incentivises cost efficiency as a means of improving nonfinancial performance, the objective function of Equation 1 is revisited and extended as in Equation 2.

$$NFP = f(e, r, p, s, t) \quad (2)$$

Where: NFP is a measure of nonfinancial performance; e is a measure of Trust cost efficiency; r is a measure of reported financial performance; s is a measure of Trust size; p is an indicator variable for PMS identity; t is an indicator variable for the year in which performance is measured.

Trust cost efficiency, e, is measured using Trust reference cost indices (*REF*) which are independent of the two performance measurement systems and which are adjusted for both case mix and market factors affecting cost levels.

Reported financial performance is measured in four ways. First, a continuous variable of reported financial performance scaled by total revenue (*FP*) is adopted. However, reported financial performance is subject to noise from transactions such as asset sales and the public capital dividend. So, to test the robustness of the results obtained using *FP*, operating surplus scaled by total revenue (*OS*) is used as a second measure of financial performance. Thirdly, underlying financial performance (*UFP*), defined as reported financial performance minus financial support, reflects an adjustment for the flexibility offered by financial support in the reporting of financial performance. Finally, for non-parametric analysis, the three-way categorical measure of reported financial performance used in the Star ratings is adopted and applied to reported financial performance in the AHC, where there was no distinction between large and small deficits. The three categories are breakeven or better; a small deficit of up to 1% of revenue; and a large deficit of more than 1% of revenue.

Size is measured as the log of Trust total revenue (*SIZE*) and, as more complex clinical cases tend to be treated in large regional centres the log of Trust total revenue represents an additional control variable for variations in case mix and complexity.

Finally, a measure of nonfinancial performance, *NFP*, which is uniform across both the Star ratings and the AHC, is constructed. This measure represents a key feature of this thesis, representing a broader and more objective measure of nonfinancial performance than has been possible in previous studies. *NFP* exploits common features of both performance measurement systems, specifically the key Government measures and the three point categorical system. Performance against the limited number of key Government measures, (Department of Health, 1999, 2000), forms the basis of *NFP,* performance against which is measured using the scoring system of the AHC which awards three points for achievement, two points for underachievement and zero for failure. The positive scoring system of the AHC is adopted as, in contrast to the Star ratings, a system of scaling already exists in the AHC for Trusts which are not subject to all measures (see Figure 6 Section 4.4.3.1,The Annual Health Check)[47]. This scoring system is applied to the key Government measures in the Star ratings so that a comparable sum score is obtained for all Trusts across the study period. Then, as the number of measures varies by PMS, by year and by Trust[48] this sum score is normalized to the equivalent of 12 measures (the maximum number in the AHC) as follows. In 2003 and 2004 when the number of key government nonfinancial measures was 8 the factor applied to the aggregated performance score was 12/8. In 2005 when there were seven key measures the factor applied was 12/7. There were ten trust observations out of a total of 441 in the Star ratings period for which a rating was produced for five to seven measures and the factor applied to these was adjusted accordingly. In the AHC all general (non-specialist) Trusts had between 10 and 12 measures. A Trust measured against 10 indicators would have its score inflated by a factor of 12/10. So a Trust which scored 30 points over 10 indicators (the maximum score) would have an *NFP* score of 36, the maximum *NFP* score. Similarly a Trust scoring 20 points over 10 indicators would have an *NFP* score of 24. The impact of this alignment of scores for Trusts measured against different

[47] The Star Ratings were based on penalty points of zero for achievement, two for underachievement and six for significant underachievement. This compares with positive scoring in the AHC of three points for achievement, two for underachievement and zero for failure.

[48] There are eight Government imposed nonfinancial measures in the Star Ratings (with the exception of 2005 when the Improving Working Lives measure was dropped) compared with 12 in the AHC. In the Star ratings there is minimal variation in the number of measures applied to each individual Trust.

numbers of measures has the effect of giving greater dispersion to the *NFP* score for the Stars system. The resulting *NFP* score has a potential range of scores of zero to 36.

All data were accessed from publicly available sources. Data on performance against PMS measures was accessed from the websites of the Commission for Health Improvement and its successor the Healthcare Commission; data on Trust reference cost indices were obtained from the Department of Health and data on reported financial performance and operating surpluses were obtained from the Laing and Buisson database of NHS financial performance.

The sample selected includes only English NHS acute hospital Trusts as hospitals in Scotland, Wales and Northern Ireland operate under different regulatory regimes. It is restricted to acute hospital Trusts and does not include mental health Trusts, primary care or community hospital services whose performance was subject to a different set of performance measures. The sample excludes acute hospital Foundation Trusts which, although subject to the same nonfinancial measures as acute NHS Trusts, were subject to a different regime of financial regulation and, in particular, are not bound by a statutory duty to break-even. As specialist hospitals have been identified in previous studies as having significantly lower efficiency than general hospitals, the sample is further sub-divided into general and specialist Trusts, to identify whether there are any significant differences. Finally, it is arguable that policy makers are most interested in the impact of performance measurement systems on poor performers, those in the tail of the performance comet, and whether, for these Trusts, manipulation of the system rather than the improvement of cost efficiency is incentivised. A further subsample therefore excludes the best performing Trusts, which are identified using the proxy of those Trusts which converted to Foundation Trust status during the study period.

5.3.5 Non-parametric analysis

Both the production possibility frontier analysis and the performance comet predict that PMS design will impact the extent to which financial performance is sacrificed in order to improve overall performance ratings. As a precursor to regression analysis, non-parametric chi-squared tests are conducted to provide

further insight into the impact of the incentive to trade financial for nonfinancial performance. The aim of these chi square tests is to establish the extent to which failure to break even in each PMS was greater than failures on other measures.

5.3.6 Modelling

To facilitate the investigation into the relationship between *NFP* and cost efficiency, *NFP* is modelled as a linear function of Trust reference cost indices (*REF*) and financial performance. In the first instance, reported financial performance (*FP*) is used, but to test for the robustness of the findings, operating surplus (*OS*) is subsequently substituted. Variations in case mix are accommodated in the Trust reference cost index (*REF*) and by the *SIZE* control variable. Finally, variation in nonfinancial performance standards across the early years of the study, and other issues such as increasing levels of NHS funding, are accommodated by the inclusion of an indicator variable, *YEAR*, coded one to six for each of the years 2002-03 to 2007-08. The following pooled, cross-sectional OLS regression is then applied, first to the whole study period (2003-2008) and then, to identify differences between the Star ratings and the AHC, to the separate periods of 2003-2005 and 2006-2008.

$$NFP_{it} = \alpha_0 + \alpha_1 REF_{it} + \alpha_2 FP_{it} + \alpha_3 SIZE_{it} + \sum_{l=m}^{n} \alpha_l YEAR_i + e_{it} \qquad (3)$$

Where: NFP_{it} is a measure of nonfinancial performance based on performance against key Government measures for Trust i in each year t of the study; REF is the Trust reference cost index, a measure of cost efficiency; FP is reported financial performance scaled by total revenue; $SIZE$ is the natural log of Trust total revenue; $YEAR$ is an indicator variable which is coded 1 to 6 for the years 2003-2008 for the whole study period; and 1 to 3 for each of the periods 2003-2005 (the Star ratings) and 2006-2008 (the AHC); e is an error term.

Finally, to test the robustness of the results from Equation 3 operating surplus is substituted for reported financial performance and the following regression applied to the same samples:

$$NFP_{it} = \alpha_0 + \alpha_1 REF_{it} + \alpha_2 OS_{it} + \alpha_3 SIZE_{it} + \sum_{l=m}^{n} \alpha_l YEAR_i + e_{it} \quad (4)$$

Where: *OS* is operating surplus scaled by total revenue.

For the next stage in the analysis interaction effects are introduced in order to identify whether any differences between the Star ratings and the AHC observed from the above regressions are significant. Interactions are effected through the introduction of an indicator variable (*PMS*) coded 1 for the Star ratings and 0 for the AHC. The following OLS regression is applied:

$$NFP_{it} = \alpha_0 + \alpha_1 REF + \alpha_2 PMS * REF_{it} + \alpha_3 FP_{it} + \alpha_4 PMS * FP_{it} + \alpha_5 SIZE_{it} + \sum_{l=2003}^{2007} \alpha_l YEAR_i + e_{it} \quad (5)$$

Where: PMS is an indicator variable coded 1 for the Star ratings (2003-2005) and zero for the AHC (2006-2008); $PMS * REF_{it}$ is the differential impact on *REF* of the Star ratings as compared with the AHC*;* $PMS * FP_{it}$ is the differential impact on *FP* of the Star ratings as compared with the AHC.

As for the individual regressions, operating surplus (*OS*) is then substituted for reported financial performance (*FP*) and the regressions re-run. Finally, the impact of financial support on the relationship between *NFP* and *REF* is tested by substituting underlying financial performance (*UFP*), defined as reported financial performance minus financial support received, for reported financial performance (*FP*).

5.4 STUDY 2: THE PERFORMANCE IMPACT OF FINANCIAL SUPPORT.

This study develops the theme of the manipulation of reported financial performance to investigate the performance and accountability consequences of financial support and to identify whether such manipulation can be in the interests of stakeholders.

5.4.1 Introduction

The system of financial support involved the shifting of income from NHS organisations with surplus funds to those in financial difficulty. Financial support often served to advance the recognition of revenue for Trusts in financial difficulty and was an approved variation from conventional UK GAAP. In contrast with the empirical literature which largely views the manipulation of performance measurement systems as opportunistic, the analytical literature on the manipulation of financial performance has explored more fully the question as to whether manipulation can be beneficial to stakeholders, for example, by countering the incentive to engage in more damaging responses to the achievement of financial objectives (Demski, Frimor and Sappington, 2004; Demski, 1998; Demski and Sappington, 1990; Scott, 2009). This study investigates the question as to whether the system of financial support generated benefits, as was intended, in the form of better performance.

In terms of the current period's performance, the provision of additional funds in the form of financial support should lead to better service standards, particularly as its provision is conditional upon a credible recovery plan. In these circumstances the additional funds contribute towards the maintenance of service standards whilst protecting the Trusts against the adverse consequences of reporting a deficit. Further, financial support could beneficially impact the future service standards of Trusts receiving it. This could occur through inertial delays in translating the receipt of financial support into service initiatives but, more significantly, could occur because the receipt of financial support allows Trusts time to implement initiatives which improve their processes and which take time to be reflected in service standards. Financial support was generally provided on a non-recurrent basis and Trusts were expected to address their financial difficulties such that they restore their ability to achieve financial balance, without further assistance. A final aspect of the provision of financial support is thus that Trusts receiving it

should demonstrate improved underlying financial performance in the year following receipt.

The alternative view of financial support, however, was that the system of financial support system was being used opportunistically to disguise poor financial performance without any accompanying improvement in service standards. In the context of a complex layering of accountability and an evolving incentive framework, concerns grew that the selection criterion of a credible recovery plan was being suspended such that 'poor' managers were given access to financial support. These managers may not have the skills to implement the necessary actions for a significant improvement in performance to be observed and may be similar to those managers in the failing Trusts identified by the Audit Commission (2006b) as becoming dependent, year after year, on financial support.

Three empirical questions are investigated in Study 2. First, whether financial support led to an improvement in service standards in the year of its receipt; second, whether the benefits accruing to the receipt of financial support continued to flow through into the service standards of the following year and third, whether, in line with its policy aims, financial support allowed time for Trusts to address their financial problems, thus leading to an improvement in underlying financial performance in the following year.

Research question 1: Did the receipt of financial support improve the contemporaneous level of nonfinancial performance?

Research question 2: Was the subsequent change in nonfinancial performance positively associated with the receipt of financial support?

Research question 3: Was there a positive association between receipt of financial support and the subsequent change in underlying financial performance?

5.4.2 Research question 1: Did the receipt of financial support improve the contemporaneous level of nonfinancial performance?

The research design draws on the objective function of equation (2)

$$NFP = f(e, r, p, s, t)$$

Where: NFP is a measure of nonfinancial performance; e is a measure of Trust cost efficiency; r is a measure of financial performance; s is a measure of Trust size; t is an indicator variable for the year in which performance is measured.

As in Study 1, NFP is a measure of nonfinancial performance constructed by reference to performance against Government service standards; Trust cost efficiency e is measured by reference to the Trust reference cost index and Trust size s is measured as the natural log of Trust total revenue. In this study, however, underlying financial performance, *UFP*), defined as reported financial performance minus financial support, is substituted for reported financial performance. This amended objective function (equation 6) provides the basis for modelling nonfinancial performance as a function of cost efficiency and underlying financial performance controlling for year and size effects. The following pooled cross-sectional OLS regression is estimated for the period, 2004-2006, when financial support was disclosed in Trust financial statements.

$$NFP_{ijt} = \alpha_0 + \alpha_1 REF_{ijt} + \alpha_2 UFP_{ijt} + \alpha_3 SIZE_{ijt} + \sum_{l=2005}^{2006} \alpha_l YEAR_l + e_{ijt} \quad (6)$$

Where: NFP_{ijt} is a measure of nonfinancial performance representing service standards for Trust i in year t for the reference group j; REF is a measure of Trust cost efficiency; UFP is underlying financial performance, which in the case of Trusts not in receipt of financial support is equal to reported financial performance; $SIZE$ is a control variable measured as the log of total assets; $YEAR$ is an indicator variable coded 1 to 3 for the years 2003-04 to 2005-2006 and e is an error term.

In the first stage of the investigation this regression is applied to a benchmark group of Trusts unaffected by financial support (reference group j). As

financial support was applied selectively to Trusts in financial difficulty this benchmark group is identified as being Trusts with reported deficits not in receipt of financial support. The regression is estimated for these Trusts in order to obtain a benchmark level of nonfinancial performance for a given level of underlying financial performance.

Coefficients from equation (6) are obtained and these are then applied to Trusts receiving financial support in order to generate a performance adjusted benchmark level of nonfinancial performance which the Trust should achieve based on its underlying financial performance. This benchmark level of nonfinancial performance is then compared with actual performance to establish whether actual performance is significantly different. The regression coefficients $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ from equation (6) and year effects are then applied to the explanatory variables of Trusts (k) receiving financial support, all of which had an underlying deficit, to obtain a benchmark level of nonfinancial performance.

$$\hat{NFP}_{ikt} = \hat{\alpha}_0 + \hat{\alpha}_1 REF_{ikt} + \hat{\alpha}_2 UFP_{ikt} + \hat{\alpha}_3 SIZE_{ikt} + \sum_{l=2005}^{2006} \hat{\alpha}_l YEAR_l \qquad (7)$$

where: UFP_{ikt} is the underlying financial performance of Trust k in year t measured as reported financial performance less financial support.

The residual from equation (8) then provides a measure of whether actual nonfinancial performance is better or worse than the benchmark level.

$$NFP_{ikt} - \hat{NFP}_{ikt} = resNFP_{ikt} \qquad (8)$$

For the purposes of investigating research question 1, whether financial support leads to an improvement in contemporaneous *NFP*, the null hypothesis of no difference between actual performance and benchmark performance is then tested:

Hypothesis 1: Ceteris paribus, $resNFP$ equals zero.

5.4.3 Research question 2: Was the subsequent change in nonfinancial performance positively associated with the receipt of financial support?

The second research question investigates whether benefits to *NFP* associated with the receipt of financial support continue to flow through into the following period.

Hypothesis 2: Ceteris paribus, there is no association between a Trust's subsequent change in nonfinancial performance and the receipt of financial support.

Equation (9), adapts the objective function from equation (6) to investigate the change in nonfinancial performance as a function of financial support controlling for changes in the other explanatory variables: cost efficiency and financial performance, size and year effects:

$$\Delta NFP_{kt} = \alpha_0 + \alpha_1 FS_{k(t-1)} + \alpha_2 \Delta REF_{kt} + \alpha_3 \Delta UFP_{kt} + \alpha_4 \Delta SIZE_{kt} + \sum_{l=2005}^{2007} \alpha_l YEAR + e_{kt} \quad (9)$$

Where: $\Delta NFP_{kt} = NFP_{kt} - NFP_{kt-1}$ is nonfinancial performance for Trust k in year t less nonfinancial performance in year $t-1$; $\Delta REF_{kt} = REF_{kt} - REF_{kt-1}$ is the change in Trust cost efficiency; $\Delta UFP_{kt} = UFP_{kt} - UFP_{kt-1}$ is the change in underlying financial performance; $\Delta SIZE_{kt} = SIZE_{kt} - SIZE_{k(t-1)}$ is the change in the log of Trust total assets; $YEAR$ is an indicator variable coded 1 to 4 for the years 2004-2007 and e is an error term.

5.4.4 Research question 3: Was there a positive association between receipt of financial support and the subsequent change in underlying financial performance?

The impact of financial support on underlying financial performance in the year of receipt is unequivocal: it serves to reduce, and often eliminate, an underlying deficit and can be measured with precision. The third research question however focuses on the proposition that financial support allows trusts time to address underlying financial problems by investigating the association between the receipt of financial support and Trusts' subsequent changes in underlying financial performance:

Hypothesis 3: Ceteris paribus, there is no association between financial support and the Trust's subsequent change in underlying financial performance.

Underlying financial performance (*UFP*) is modelled (Model 3) as a function of financial support (*FS*) controlling for changes in nonfinancial performance (*NFP*), cost efficiency (*REF*), size and year effects. The following OLS regression is estimated:

$$\Delta UFP_{kt} = \alpha_0 + \alpha_1 FS_{k(t-1)} + \alpha_2 \Delta REF_{kt} + \alpha_3 \Delta NFP_{kt} + \alpha_4 \Delta SIZE_{kt} + \sum_{l=2005}^{2007} \alpha_l YEAR + e_{kt} \quad (10)$$

5.5 SUMMARY OF RESEARCH DESIGN

This Chapter has developed the research questions to be investigated in this thesis and the models to be used to facilitate the investigations.

Study 1 focuses on a key policy objective of PMS design in the NHS, the incentivisation of cost efficiency as a means of improving service standards. The availability of data on Trust cost efficiency and on nonfinancial performance facilitates the investigation of the relationship between cost efficiency and service standards. This question is a key concern of the literature on cost efficiency because of the potential for trading off service standards in the pursuit of cost efficiency. Further, the replacement of the Star ratings system by the Annual Health Check in 2006 provides the conditions for a natural experiment which compares the extent to which the AHC was better at incentivising cost efficiency than the Star ratings. The prediction that the AHC will be more effective than the Star ratings in incentivising cost efficiency is predicated on an analysis of the two performance systems and the identification that there is greater scope for manipulation of the Star ratings system than the AHC. A key aspect of manipulation that is identified is the manipulation of financial performance. The potential for this form of manipulation is more exaggerated in the Stars, not only because there are stronger short run incentives for going into deficit, but also because, throughout the whole period of the Star ratings, the policy of financial support was in operation.

In contrast, financial support was only available to Trusts in financial difficulty in the first year of the AHC, after which the policy was abandoned. Study 2 focuses

on the manipulation of reported financial performance arising from the accounting policy adopted for financial support and investigates whether such manipulation, which represents a form of income shifting and which in the majority of cases serves to advance the recognition of revenue in Trusts receiving financial support, can have beneficial consequences for stakeholders. Specifically this study investigates first, whether financial support leads to an improvement in nonfinancial performance (*NFP*) in the year of its receipt; second, whether any benefits to *NFP* flow through into the subsequent period; and thirdly, whether, consistent with its policy aims, financial support allows a Trust time to address underlying financial difficulties with the result that underlying financial performance improves in the subsequent period.

To facilitate these two studies, an objective function which draws on Government policy as encapsulated in the NHS Plan (Department of Health, 2000) and the Performance Assessment Framework (Department of Health, 1999) is generated. This objective function, which characterises nonfinancial performance in the presence of a financial resource constraint as a function of cost efficiency, is applied in the context of production possibility frontier analysis to two stylised hospital scenarios in order to generate predictions about the impact of PMS design on the relationship between nonfinancial performance, cost efficiency and reported financial performance. This analysis is applied to the Star ratings and the AHC to derive a prediction that the AHC will be more effective at incentivising cost efficiency as a means of improving service standards than the Star ratings. Common characteristics of the Star ratings and the AHC are then exploited in order to construct a uniform measure of nonfinancial performance, based on key Government measures, across both PMS for the whole study period.

The availability of data on Trust cost efficiency, in the form of the reference cost index and the availability of data on nonfinancial performance which facilitates the construction of a uniform measure of nonfinancial performance across all Trusts for the period of study, facilitates the investigation through non parametric and regression analysis of the main research questions in this thesis: first: the relationship between service standards and cost efficiency and whether the AHC was more effective at incentivising cost efficiency as a means of improving service standards than the Star ratings, and second: whether benefits accrued to stakeholders as a consequence of the policy of financial support.

CHAPTER 6 FINDINGS

This Chapter commences with a review of the characteristics of NHS acute hospital Trusts as revealed by descriptive statistics. It continues with the findings from the two studies: Study 1, an investigation of the relationship between service standards and cost efficiency incorporating the relative impact of the Star ratings and the AHC; and Study 2, an investigation of the performance impact of financial support, a form of income shifting between NHS organisations, whose accounting policy served to manipulate reported financial performance and undermine public accountability.

6.1 DESCRIPTIVE STATISTICS

Descriptive statistics illustrating the economic significance of NHS acute hospitals in terms of reported total revenue during the study period are reported in Table 1.

Panel A show that the full sample of Trusts (854 Trust years) reported mean revenue of £191m representing total revenue of more than £160bn over the study period. Mean revenue in 2008 was 67% higher than in 2003 and the largest Trust had revenue of £838m in 2008. The mean reported revenue during the period of the Star ratings was £170m as compared with £220m for the AHC. The number of observations in the sample declines steadily over the period from 176 in 2003 to 100 in 2008 as a result of mergers and, from 2004-05, from Trusts converting to Foundation Trust status. Foundation Trusts, being subject to a different regulatory regime, have been excluded from the sample. As specialist Trusts were identified as a special category of acute Trust in the Star ratings system (see, for example, Healthcare Commission, 2005b) and as specialist Trusts have been identified by prior frontier analysis research as having significantly lower efficiency than general Trusts, the sample is subdivided into general and specialist Trusts in Panels B and C respectively. The identity of the specialist Trusts is determined in accordance with the Healthcare Commission's classification (Healthcare Commission, 2005). Panel B shows that the mean reported revenue of general Trusts (769 observations) was somewhat bigger at c.202m than the mean for the full sample and Panel C shows that the mean reported revenue of specialist Trusts (85 observations) was much lower at £87m.

Table 1: Trust Total Revenue 2003-2008

Panel A: Full sample

Year	No. obs	Mean	Std. Devn	Min.	Max	Median
		£000	£000	£000	£000	£000
2003	176	156,103	96,048	11,198	596,678	125,602
2004	173	170,710	104,591	11,523	627,148	138,895
2005	148	186,083	108,684	12,981	677,981	156,772
2006	140	200,914	118,671	42,855	721,415	164,455
2007	117	217,468	129,787	47,868	757,446	177,619
2008	100	250,003	155,863	67,043	838,148	194,297
2003-2008	**854**	**191,006**	**119,993**	**11,198**	**838,148**	**157,992**
Star ratings	497	170,115	103,417	11,198	677,981	139,446
AHC	357	220,089	134,690	42,855	838,148	177,619

Panel B: General acute Trusts

Year	No. obs	Mean	Std. Devn	Min.	Max	Median
		£000	£000	£000	£000	£000
2003	157	166,974	95,124	51,114	596,678	133,500
2004	154	183,025	103,357	56,286	627,148	147,409
2005	132	198,679	107,150	64,201	677,981	164,863
2006	126	212,412	118,324	68,162	721,415	172,997
2007	107	227,119	130,176	70,300	757,446	182,029
2008	93	258,486	157,036	73,006	838,148	196,207
2003-2008	**769**	**202,511**	**119,637**	**51,114**	**838,148**	**165,053**
Star ratings	443	182,001	102,264	51,114	677,981	148,465
AHC	326	230,383	135,082	68,162	838,148	182,218

Panel C: Specialist Trusts

Year	No. obs	Mean	Std. Devn	Min.	Max	Median
		£000	£000	£000	£000	£000
2003	19	66,273	41,463	11,198	163,350	56,762
2004	19	70,899	43,742	11,523	176,682	60,532
2005	16	82,169	50,760	12,981	196,007	64,804
2006	14	97,427	56,051	42,855	220,075	74,421
2007	10	114,197	67,581	47,868	247,048	82,022
2008	7	137,288	82,338	67,043	270,693	89,371
2003-2008	**85**	**86,917**	**56,351**	**11,198**	**270,693**	**65,959**
Star ratings	54	72,610	44,788	11,198	196,007	60,223
AHC	31	111,838	65,858	42,855	270,693	81,122

Total revenue reported in acute hospital NHS Trust financial statements
Source: Laing and Buisson
Star ratings: 2003-2005
AHC: Annual Health Check 2006-2008

Tables 2-4 provide descriptive statistics for each of the main variables*, NFP,* a measure of nonfinancial performance (Table 2); *REF*, a measure of Trust cost efficiency (Table 3) and *FP,* reported financial performance scaled by total revenue (Table 4). In each case results are reported for the full sample (Panel A, 854 observations), general Trusts (Panel B, 769 observations) and specialist Trusts (Panel C, 85 observations).

In Table 2 Trust nonfinancial performance (*NFP*) is measured with respect to performance against key Government measures and has a maximum value of 36. The minimum observed level of *NFP* is 5, recorded by a specialist Trust. The minimum level of *NFP* for a general Trust (Panel B) is 17. The mean of 33.2 for the full sample (Panel A) is consistent with the clustering predicted by the production possibility frontier analysis and the performance comet. This clustering is predicted to arise from efficient Trusts operating on the frontier, those which transfer resources out of unmeasured activities and those that trade financial for nonfinancial performance (Figure 14). The difference between specialist and general NFP scores is minimal, with a mean for specialist Trusts of 33.12 compared with 33.18 for general Trusts. The higher mean and median, in Panels A and B, for the Star Ratings (34 and 36 respectively) as compared with the Annual Health Check (32 and 32) are consistent with the narrower range of measures and the greater scope for manipulation in the Star ratings. The higher standard deviation of 3.6 in the AHC as compared with 3.1 in the Star ratings also points to the AHC being more discriminating of performance.

To determine whether the distribution of *NFP* is different across PMS a test of the null hypothesis of no difference in the distribution of nonfinancial performance, *NFP*, in the Star ratings compared to the AHC using the Kolmogorov-Smirnov statistic (not tabulated) is rejected ($D = 0.481$; $p = 0.000$).

Table 2: Nonfinancial performance

Panel A: Full sample

		NFP				
Year	No. obs	Mean	Std. Devn.	Min.	Max	Median
2003	176	34.26	2.95	21	36	36
2004	173	34.54	2.75	17	36	36
2005	148	33.95	3.62	5	36	34
2006	140	31.50	3.28	20	36	32
2007	117	32.28	3.64	12	36	33
2008	100	31.14	3.88	17	36	32
2003-2008	854	33.17	3.56	5	36	34
Star ratings	497	34.26	3.11	5	36	36
AHC	357	31.66	3.60	12	36	32

Panel B: General Trusts

		NFP				
Year	No. obs	Mean	Std. Devn.	Min.	Max	Median
2003	157	34.15	3.04	21	36	36
2004	154	34.50	2.75	17	36	36
2005	132	33.98	2.81	21	36	34
2006	126	31.51	3.26	20	36	32
2007	107	32.44	3.17	22	36	33
2008	93	31.33	3.52	17	36	32
2003-2008	769	33.18	3.31	17	36	34
Star ratings	443	34.22	2.88	17	36	36
AHC	326	31.76	3.33	17	36	32

Panel C: Specialist Trusts

		NFP				
Year	No. obs	Mean	Std. Devn.	Min.	Max	Median
2003	19	35.11	1.97	30	36	36
2004	19	34.89	2.81	24	36	36
2005	16	33.69	7.69	5	36	36
2006	14	31.43	3.57	22	36	31
2007	10	30.60	7.01	12	36	33
2008	7	28.57	7.09	19	36	31
2003-2008	85	33.12	5.34	5	36	36
Star ratings	54	34.59	4.63	4.8	36	36
AHC	31	30.52	5.62	12	36	31

NFP : Trust nonfinancial performance measured against key Government measures on a scale of zero to 36: Source: Healthcare Commission/Commission for Health Improvement.

Star ratings: 2003-2005; AHC: Annual Health Check 2006-2008

Descriptive statistics for *REF*, an inverse measure of relative Trust cost efficiency, are shown in Table 3. The mean reference cost over the study period was 100.8, with a standard deviation of 10.9 and a range of 75 – 162. However there are notable differences between the reference costs of non-specialist Trusts (Panel B) and specialist Trusts (Panel C). The mean reference cost of non-specialist Trusts is 98.9, with a minimum of 78 and a maximum of 138. Both the mean and the range are smaller than for specialist Trusts which have a mean of 118 and a range of 75-162. The Kolmogorov-Smirnov statistic (not tabulated) of D = 0.054 (p = 0.000) indicates that there is no difference in the distribution of Trust reference cost indices as between the Star ratings and the AHC.

Descriptive statistics for reported financial performance as a percentage of revenue are shown in Table 4. With the exception of 2008, a mean deficit was reported in each year of the study period (Panel A). This mean deficit increased in the period from 2002-03 to 2005-06, when it reached a maximum of 2.6% of revenue, thereafter steadily reducing with a mean surplus of 0.9% of revenue being recorded in 2008. The maximum deficit of 28% of revenue was recorded in 2006, this being a general Trust (Panel B). Consistent with the statutory duty to breakeven the median financial performance was a small surplus, for both specialist and general Trusts. The overall mean for specialist Trusts is however positive at 0.17% of revenue as compared with a mean deficit for general Trusts of -0.95%. A test of the null hypothesis of no difference in the distribution of financial performance in the Star Ratings compared to the AHC (Panel A) using the Kolmogorov-Smirnov statistic (not tabulated) is rejected (D = 0.371; p = 0.000).

Table 3: Cost efficiency

Panel A: Full sample

		REF				
Year	No. obs	Mean	Std. Devn.	Min.	Max	Median
2003	176	100.4	11.0	78.5	159.1	99.4
2004	173	100.5	10.2	79.2	146.7	99.4
2005	148	100.2	10.1	75.2	149.0	99.4
2006	140	101.2	10.1	83.0	153.4	100.1
2007	117	101.9	12.7	80.7	158.1	100.9
2008	100	101.1	12.1	81.2	161.9	99.7
2003-2008	854	100.8	10.9	75.2	161.9	99.7
Star ratings	497	100.4	10.4	75.2	159.1	99.4
AHC	357	101.4	11.6	80.7	161.9	100.2

Panel B: General Trusts

		REF				
Year	No. obs	Mean	Std. Devn.	Min.	Max	Median
2003	157	98.6	7.9	78.5	126.2	98.6
2004	154	98.5	7.2	79.2	118.4	98.5
2005	132	98.6	7.5	77.8	115.8	99.0
2006	126	99.4	7.0	83.0	119.3	99.6
2007	107	99.7	9.5	80.7	138.0	100.2
2008	93	98.8	7.2	81.2	114.7	98.3
2003-2008	769	98.9	7.7	77.8	138.0	98.9
2003-2005	443	98.6	7.5	77.8	126.2	98.6
2006-2008	326	99.3	8.0	80.7	138.0	99.6

Panel C: Specialist Trusts

		REF				
Year	No. obs	Mean	Std. Devn.	Min.	Max	Median
2003	19	115.1	19.3	82.2	159.1	110.3
2004	19	116.4	16.2	86.3	146.7	113.7
2005	16	113.1	17.4	75.2	149.0	114.9
2006	14	117.3	17.9	87.1	153.4	115.8
2007	10	126.2	16.6	104.2	158.1	126.1
2008	7	132.1	20.8	101.6	161.9	128.8
2003-2008	85	118.1	18.2	75.2	161.9	118.2
Star ratings	54	115.0	17.4	75.2	159.1	113.5
AHC	31	123.5	18.6	87.1	161.9	121.5

REF: Trust cost efficiency measured (inversely) by the Trust reference cost index. Mean efficiency = 100. Source: Department of Health
Star ratings: 2003-2005; AHC: Annual Health Check 2006-2008

Table 4: Reported Financial Performance Scaled by Total Revenue

Panel A: Full sample

		FP				
Year	No. obs	Mean	Std. Devn.	Min.	Max	Median
2003	176	-0.298	2.440	-20.481	6.596	0.007
2004	173	-0.547	1.485	-7.517	3.026	0.004
2005	148	-1.382	3.157	-19.934	6.501	0.002
2006	140	-2.589	4.959	-28.027	4.367	0.003
2007	117	-0.794	3.343	-18.456	4.920	0.073
2008	100	0.929	3.279	-21.967	7.998	1.075
2003-2008	854	-0.836	3.340	-28.027	7.998	0.011
Star ratings	497	-0.708	2.454	-20.481	6.596	0.004
AHC	357	-1.015	4.276	-28.027	7.998	0.073

Panel B: General Trusts

		FP				
Year	No. obs	Mean	Std. Devn.	Min.	Max	Median
2003	157	-0.380	2.513	-20.481	5.738	0.007
2004	154	-0.580	1.507	-7.517	3.026	0.003
2005	132	-1.498	3.278	-19.934	6.501	0.001
2006	126	-2.931	5.110	-28.027	4.367	0.000
2007	107	-0.915	3.433	-18.456	4.920	0.069
2008	93	0.919	3.392	-21.967	7.998	1.071
2003-2008	769	-0.947	3.469	-28.027	7.998	0.010
Star ratings	443	-0.783	2.535	-20.481	6.501	0.004
AHC	326	-1.171	4.429	-28.027	7.998	0.056

Panel C: Specialist Trusts

		FP				
Year	No. obs	Mean	Std. Devn.	Min.	Max	Median
2003	19	0.381	1.610	-2.069	6.596	0.020
2004	19	-0.280	1.291	-5.514	0.785	0.004
2005	16	-0.430	1.651	-6.190	1.390	0.009
2006	14	0.497	0.682	-0.650	1.705	0.376
2007	10	0.493	1.794	-3.593	2.817	0.499
2008	7	1.065	0.980	0.030	2.570	1.319
2003-2008	85	0.169	1.451	-6.190	6.596	0.023
Star ratings	54	-0.092	1.532	-6.190	6.596	0.009
AHC	31	0.624	1.191	-3.593	2.817	0.397

FP = surplus/deficit reported in acute hospital NHS Trust financial statements scaled by total revenue expressed as a percentage. Source: Laing and Buisson
Star ratings: 2003=2005; AHC: Annual Health Check 2006-2008

Finally, as notable differences between general and specialist Trusts are observed in Tables 1-4, two tailed tests of means are performed to establish whether the difference in means is significant. The results of these t-tests are shown in Table 5. The mean revenue, cost efficiency and reported financial performance for specialist Trusts are all significantly different from those of general Trusts. Consistent with their narrower range of services, specialist Trusts were significantly smaller, (mean revenue £87m as compared with £202m, Panel A), had higher reference costs (mean *REF* 118 as compared with 99, Panel C) and better reported financial performance (+0.17% of revenue as compared with -0.95%, Panel D). There is no significant difference however in *NFP (*Panel B). Particularly notable is the significant difference in cost efficiency. The Trust reference cost index at 118 suggests that specialist Trusts are 18% less efficient than average. These findings are consistent with the findings of frontier analysis research which suggests that specialist hospitals are less efficient than general acute hospitals, perhaps because they are unable to exploit economies of scale and scope. The fact that specialist Trusts generally break-even also suggests that their higher cost base is recognised by commissioners when service level agreements are negotiated.

Table 5: Two tailed tests of means: general and specialist Trusts

Panel A: Trust total revenue

	No. Obs.	Mean	Std. Err.	Std. Dev.
		£000	£000	£000
General	769	202,511	4,314	119,637
Specialist	85	86,917	6,112	56,351
Total	854	191,006	4,106	119,993
diff		115,595	13,140	
t		8.797		
p		0.000		

Panel B: *Nonfinancial performance*

NFP	No. Obs.	Mean	Std. Err.	Std. Dev.
General	769	33.2	0.1	3.3
Specialist	85	33.1	0.6	5.3
Total	854	33.2	0.1	3.6
diff		0.1	0.4	-0.7
t		0.152		
p		0.879		

Panel C: Cost efficiency

REF	No. Obs.	Mean	Std. Err.	Std. Dev.
General	769	98.9	0.3	7.7
Specialist	85	118.1	2.0	18.2
Total	854	100.8	0.4	10.9
diff		-19.2	1.1	-21.3
t		-18.092		
p		0.000		

Panel D: Reported financial performance as a % of revenue

FP	No. Obs.	Mean	Std. Err.	Std. Dev.
General	769	-0.947	0.125	3.469
Specialist	85	0.169	0.157	1.451
Total	854	-0.836	0.114	3.340
diff		-1.116	0.380	-1.862
t		-2.938		
p		0.003		

NFP: Trust nonfinancial performance measured against key Government measures on a scale of zero to 36; *REF*: Trust reference cost index, an inverse measure of Trust cost efficiency, average efficiency =100 (Source: Department of Health); *FP*: surplus/deficit reported in acute hospital NHS Trust financial statements scaled by total revenue expressed as a percentage (Source: Laing and Buisson)

The best performing Trusts were gradually converting to Foundation Trusts from 2004-05 onwards, Table 6 provides further insight into the characteristics of these converting Trusts. To ensure a like for like comparison, between converters and non-converters specialist Trusts are excluded from the sample. Not only do they have different characteristics from the rest of the sample but proportionately more specialist Trusts converted during the study period (45% of all specialist observations) period than for the general Trusts (25% of all general observations). As compared with non-converters, Trusts which converted to Foundation Trust status during the study period were smaller (mean revenue £177m as compared with £211m) and, consistent with the requirement for generally high levels of performance, were characterised by higher nonfinancial performance (mean *NFP* 35.0 as compared with 32.6), lower reference costs (mean *REF* 96.7 as compared with 99.6) and better reported financial performance (mean -0.03% of revenue as compared with -1.25%). The differences between converters and non-converters although significant are much smaller than the differences between specialist and general Trusts.

Table 6: Two tailed tests of means – Trusts converting to Foundation Trust status

Panel A: Trust total revenue

	No. Obs.	Mean	Std. Err.	Std. Dev.
		£000	£000	£000
Non converters	578	210,904	5,265	126,574
Converters	191	177,114	6,607	91,316
Total	769	202,511	4,314	119,637
Difference		33,790	9,917	14,323
t		3.407		
p		0.000		

Panel B: *Nonfinancial performance*

NFP	No. Obs.	Mean	Std. Err.	Std. Dev.
Nonconverters	578	32.6	0.1	3.5
Converters	191	35.0	0.1	1.8
Total	769	33.2	0.1	3.3
Difference		-2.5	0.3	-3.0
t		-9.422		
p		0.000		

Panel C: Trust cost efficiency

REF	No. Obs.	Mean	Std. Err.	Std. Dev.
Nonconverters	578	99.6	0.3	7.9
Converters	191	96.7	0.5	6.8
Total	769	98.9	0.3	7.7
Difference		2.8	0.6	1.6
t		4.474		
p		0.000		

Panel D: Reported financial performance as a % of revenue

FP	No. Obs.	Mean	Std. Err.	Std. Dev.
Non converters	578	-1.252	0.163	3.920
Converters	191	-0.026	0.066	0.915
Total	769	-0.947	0.125	3.469
Difference		-1.225	0.286	-1.787
t		-4.279		
p		0.000		

NFP: Trust nonfinancial performance measured against key Government measures on a scale of zero to 36, (Source: Healthcare Commission); *REF*: Trust reference cost index, an inverse measure of Trust cost efficiency, average efficiency =100 (Source: Department of Health); *FP*: surplus/deficit reported in acute hospital NHS Trust financial statements scaled by total revenue expressed as a percentage (Source: Laing and Buisson)

The above descriptive analysis of the full sample of NHS acute Trusts provides the basis for an investigation of the impact of the Star ratings and the AHC on the incentivisation of cost efficiency as a means of improving service standards. First, the full sample which includes specialist Trusts is investigated but as specialist Trusts exhibit very different characteristics from general Trusts, an important sub-sample for investigation is that of general Trusts only . Further, as a robustness test, for the impact of PMS design on weaker Trusts, a further sub-sample which excludes, as a proxy for the best performing Trusts, those which converted to Foundation Trust status during the period of study, is also investigated.

6.2 FINDINGS STUDY 1: PERFORMANCE MEASUREMENT SYSTEM DESIGN AND THE INCENTIVISATION OF COST EFFICIENCY

6.2.1 Non parametric analysis

As a precursor to multivariate analysis insight into the extent to which Trusts responded to the incentives to incur a deficit as a short term tactic for enhancing their performance rating is considered for each PMS. Using the Star ratings categorisation of reported financial performance of breakeven or more, a small deficit of up to 1% of revenue and a large deficit of more than 1% of revenue, a categorical analysis of reported financial performance and *NFP* is reported for the Star ratings and the AHC in Table 7, Panels A and B respectively. To ensure a like-for-like comparison those Trusts which converted to Foundation Trust status during the period of study are excluded from the sample. The result is a total sample of 625 observations.

Panel A shows that, in the Star ratings, consistent with a clustering predicted by the production possibility frontier analysis and the performance comet (Figure 14), 42% (134/318) of Trusts achieve the maximum *NFP* score of 36. Of these, 85 (64%) reported break-even or better. A further 49 Trusts (36%) achieve *NFP*=36 but record a deficit. In 32 cases this deficit was a large deficit which, for a Trust with average revenue, would amount to about £2m. The very tight clustering of *NFP* scores at NFP=35 and NFP=36 is consistent with the performance comet analysis (Figure 14, Section 4.4.3.1) and the incentives identified in Figure 12. (Section 1.1.1). Panel B shows that the AHC was more discriminating of performance than the Star ratings with a reduced clustering of *NFP* scores. This wider dispersion is consistent with the lower scope for focusing on a narrow range of indicators and the greater emphasis on financial performance which offers less scope for improving a Trust's overall performance rating by incurring a deficit to improve *NFP*.

Table 7 also shows that the proportion of Trusts achieving breakeven or better increased from 58% in the Star ratings to 68% in the AHC and that this was achieved by a reduction in small deficits from 12% to 2%. The proportion of Trusts reporting a large deficit remained largely unchanged (30% in the Star ratings, 29% in the AHC). These findings are consistent with the much heavier penalties in the AHC for incurring a deficit which, prima facie, appears to have reduced the incentive

to incur a small deficit. The unchanged proportion of large deficits is consistent with no further penalty in the AHC once a deficit is incurred.

Table 8 shows the same categorical analysis but excluding specialist trusts as well as Trusts converting to Foundation status. This results in a total number of observations of 578. Specialist Trusts are identified by reference to the Healthcare Commission's classification of acute Trusts as can be found for example in the 2004-05 Star ratings (Healthcare Commission, 2005). The results are very similar to those in Table 7: the percentage of Trusts with large deficits (> 1% revenue) in both the Star ratings and the Annual Health is similar at 31% as compared with 30%; there is a similar reduction in the percentage of small deficits (up to 1% revenue) from 12% in the Stars to 2% in the AHC; and the percentage in Trusts achieving breakeven or better rises from 56% in the Stars (comparator 58%) to 67% in the AHC (comparator 68%). The difference in dispersion between the NFP scores in the AHC as compared with the Star ratings is also evident, with the main difference being in the proportion of Trusts achieving the highest score of 36, which falls from 42% to 40%.

Table 7: Categorical analysis of reported financial performance and nonfinancial performance excluding Trusts converting to Foundation Trust status during the study period

Panel A: Star ratings

NFP	Large deficit	Small deficit	Breakeven or	Total	% of total
	>1% revenue	<=1%revenue	more		obs.
	No. obs	No. obs	No. obs	No. obs	
5	1	0	0	1	0.31%
17	1	0	0	1	0.31%
21	4	0	0	4	1.26%
24	1	0	2	3	0.94%
26	3	1	1	5	1.57%
27	2	2	2	6	1.89%
29	5	1	5	11	3.46%
30	7	2	9	18	5.66%
31	3	0	5	8	2.52%
32	3	5	15	23	7.23%
33	13	3	12	28	8.81%
34	14	3	21	38	11.95%
35	6	5	27	38	11.95%
36	32	17	85	134	42.14%
Total	95	39	184	318	100.00%
% of obs.	29.87%	12.26%	57.86%	100.00%	

Panel B: Annual Health Check

NFP	Large deficit	Small deficit	Breakeven	Total	% of total
	>1% revenue	<=1%revenue	or more		obs.
	No. obs	No. obs	No. obs	No. obs	
12	0	0	1	1	0.33%
17	0	0	1	1	0.33%
19	0	0	1	1	0.33%
20	1	0	0	1	0.33%
21	0	0	1	1	0.33%
22	1	0	1	2	0.65%
23	3	0	4	7	2.28%
24	1	0	4	5	1.63%
25	4	0	1	5	1.63%
26	4	0	3	7	2.28%
27	4	1	4	9	2.93%
28	7	2	9	18	5.86%
29	6	1	15	22	7.17%
30	3	1	8	12	3.91%
31	10	2	25	37	12.05%
32	10	0	34	44	14.33%
33	14	1	21	36	11.73%
34	9	0	24	33	10.75%
35	7	0	36	43	14.01%
36	5	0	17	22	7.17%
Total	89	8	210	307	100.00%
% of obs.	28.99%	2.61%	68.40%	100.00%	

NFP: Trust nonfinancial performance measured against key Government measures on a scale of zero to 36.

Table 8: Categorical analysis of reported financial performance and nonfinancial performance excluding specialist Trusts and Trusts converting to Foundation Trust status

Panel A: Star ratings

NFP	Large deficit	Small deficit	Breakeven or	Total	
	>1% revenue	<=1%revenue	more		
	No. obs	No. obs	No. obs	No. obs	% total obs.
17	1	0	0	1	0.34%
21	4	0	0	4	1.36%
24	1	0	1	2	0.68%
26	3	1	1	5	1.70%
27	2	2	2	6	2.04%
29	5	1	5	11	3.74%
30	6	2	8	16	5.44%
31	3	0	5	8	2.72%
32	3	5	15	23	7.82%
33	13	3	12	28	9.52%
34	13	2	18	33	11.22%
35	6	5	27	38	12.93%
36	31	16	72	119	40.48%
Total	91	37	166	294	100.00%
% total obs.	30.95%	12.59%	56.46%	100.00%	

Panel B: Annual Health Check

NFP	Large deficit	Small deficit	Breakeven	Total	
	>1% revenue	<=1%revenue	or more		
	No. obs	No. obs	No. obs	No. obs	% total obs.
17	0	0	1	1	0.35%
20	1	0	0	1	0.35%
22	1	0	0	1	0.35%
23	3	0	4	7	2.46%
24	1	0	3	4	1.41%
25	4	0	1	5	1.76%
26	4	0	3	7	2.46%
27	4	1	4	9	3.17%
28	7	2	9	18	6.34%
29	6	0	14	20	7.04%
30	3	0	7	10	3.52%
31	10	2	23	35	12.32%
32	10	0	32	42	14.79%
33	13	1	19	33	11.62%
34	9	0	22	31	10.92%
35	7	0	35	42	14.79%
36	5	0	13	18	6.34%
Total	88	6	190	284	100.00%
% total obs.	30.99%	2.11%	66.90%	100.00%	

NFP: Trust nonfinancial performance measured against key Government measures on a scale of zero to 36

Specialist Trusts are identified by reference to the Healthcare Commission's classification of acute Trusts. Trusts converting to Foundation Trust status are identified by reference to Monitor, the Foundation Trust regulator and the Laing and Buisson database of NHS financial statements.

To provide additional insight into the impact of the incentive to trade financial for nonfinancial performance a non-parametric chi-square test using the full sample of 854 observations is provided in Table 9 (Star ratings) and Table 10 (AHC). Performance against each key Government measure is classified according to the three way classification characteristic of both systems. To allow inspection of the contribution of each individual measure to the overall chi square statistic the actual and expected frequencies for each measure is reported in addition to its contribution to the overall statistic. The full sample includes both specialist Trusts and Trusts which converted to Foundation Trust status during the study period, both of which exhibited significantly better financial performance than non-converting general Trusts. The results from this analysis will therefore present a conservative estimate of the incidence of Trusts' failure to achieve the breakeven target as compared with their failure to meet other targets.

The actual frequencies for the Star ratings reported in Table 9 reveal the very high level of achievement of all targets, both financial and nonfinancial (3701 out of 4194 observations, 88%). A similar picture for the AHC is revealed by Table 10 where the concentration of target achievement at 78% remains high (3193/4105) Consistent with the strong short run incentive in the Star ratings for a Trust to move into deficit to enhance nonfinancial performance, Table 9 shows that the largest contributor to the chi square statistic of 526 ($p = 0.000$) is the financial breakeven measure where the actual frequency of 148 Trusts reporting a deficit compares with the predicted frequency of 59. Particularly notable is the number of large deficits, 106, compared with the predicted 25. The contribution of the other (nonfinancial) measures is generally small, except for the four hour wait for accident and emergency treatment (AE4).

Table 10 shows that in the AHC the discrepancy between the actual and predicted number of deficits is much lower, with actual deficits at 101 compared with predicted deficits of 80. This reduction is mainly a consequence of the reduction in small deficits. Indeed the number of small deficits (9) is actually less than predicted (52). The incidence of large deficits remains comparable with the Star ratings at 92 actual observations compared with the predicted 28. Untabulated median tests for the Star Ratings and the AHC were also performed for financial and nonfinancial measures and the null hypotheses of no difference for both were rejected. A further observation is that, in contrast with the Star ratings, financial breakeven no longer

contributes the major part of the chi-square score. Cancelled operations (CanOps, 326) is much higher than the breakeven contribution of 188, and Thrombolysis (Thromb, 156) is comparable.[49] These findings are consistent with the prediction that the Star ratings, as compared with the AHC, incentivised the incurrence of deficits in order to enhance performance against non-financial measures as a result of lower measurement diversity and weighting given to financial measures as compared with the AHC.

The production possibility frontier analysis (Section 5.3.2) predicts that, where the relative weighting given to financial measures is low, inefficient Trusts are incentivised to go into deficit to improve their nonfinancial performance. T-tests (untabulated) of the reference costs of deficit and surplus Trusts however show no significant differences, whatever sample is selected. The deficits of Trusts with reference costs of more than 100 (i.e. greater than average) are, however, significantly greater than the deficits of Trusts with reference costs of less than or equal to 100 for samples which exclude specialist Trusts. The absence of significant differences for samples which include specialist Trusts are consistent with the descriptive statistics which show that specialist Trusts have significantly higher reference costs than the rest of the sample but that on average, in contrast with the rest of the population, they report a small surplus for the study period.

[49] The contribution of cancelled operations contribution to the chi square statistic is interesting. In the tighter performance measurement environment of the AHC, where a more balanced weighting is given to financial performance, the cancelled operations statistic is consistent with the focus of manipulation moving elsewhere. Manipulation of the cancelled operations target can serve to improve performance against waiting times, for which there is more than one target. (See, for example, Audit Commission, 2003). If such manipulation were taking place this would represent an example of a tighter financial measurement regime resulting in adverse consequences for service standards.

Table 9: Performance against key Government measures in the Star rating system 2003-2005

DERIVATION OF THE CHI-SQUARE STATISTIC

Measure	1	2	3	4	5	6	7	8	9	10	
	AE4	CAN2W	Inpatient	Outpatient	CanOps	AE12	Clean	IWL	Booking	Financial breakeven	Total
Actual frequency	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.
Failed	23	5	32	15	1	13	0	9	4	106	208
Underachieved	87	44	25	30	7	19	19	1	11	42	285
Achieved	343	418	440	452	161	415	478	339	306	349	3701
Total	453	467	497	497	169	447	497	349	321	497	4194
Predicted frequency											
Failed	22	23	25	25	8	22	25	17	16	25	208
Underachieved	31	32	34	34	11	30	34	24	22	34	285
Achieved	400	412	439	439	149	394	439	308	283	439	3701
Total	453	467	497	497	169	447	497	349	321	497	4194
Chi-square derivation											
Failed	0	14	2	4	7	4	25	4	9	268	337
Underachieved	103	5	2	0	2	4	6	22	5	2	152
Achieved	8	0	0	0	1	1	4	3	2	18	37
Total	111	19	4	5	9	9	35	29	16	289	**526**

KEY: AE4 = 4 hour accident and emergency wait; CAN2W = 2 week cancer waits; Inpatient = inpatient waiting time; Outpatient = outpatient waiting time; CanOps = number of cancelled operations (2003 only); AE12 =12 hour accident and emergency waits for admission; Clean= cleanliness of hospitals; IWL = implementation of the Improving Working Lives program (2003 and 2004 only); Booking = implementation of patient booking system (2004 and 2005 only).

Table 10: Performance against key Government measures in the Annual Health Check 2006-2008

DERIVATION OF THE CHI-SQUARE STATISTIC

Measure	1	2	3	4	5	6	7	8	9	10	11	12		Financial Break even	TOTAL
Actual Frequency	AE4	CAN2W	Inpatient	Outpatient	CanOps	CAN1M	CAN2M	THROM	REVAS	CHEST	Del.Trans	Choice			
	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.	Freq.		Freq.	Freq.
Failed	23	3	24	11	84	1	15	14	6	12	13	19		92	317
Underachieved	53	5	49	16	136	7	56	110	2	41	41	70		9	595
Achieved	257	333	284	330	136	346	268	141	95	276	203	268		256	3193
Total	333	341	357	357	356	354	339	265	103	329	257	357		357	4105
Predicted Frequency															
Failed	26	26	28	28	27	27	26	20	8	25	20	28		28	317
Underachieved	48	49	52	52	52	51	49	38	15	48	37	52		52	595
Achieved	259	265	278	278	277	275	264	206	80	256	200	278		278	3193
Total	333	341	357	357	356	354	339	265	103	329	257	357		357	4105
Chi-square derivation															
Failed	0	21	0	10	116	25	5	2	0	7	2	3		151	343
Underachieved	0	40	0	25	138	38	1	133	11	1	0	6		35	430
Achieved	0	17	0	10	72	18	0	21	3	2	0	0		2	144
Total	1	78	1	45	326	82	6	156	14	10	3	9		188	917

KEY: AE4= 4hr accident and emergency wait; CAN2W = 2 week cancer waits; Inpatient = inpatient waiting time; Outpatient = outpatient waiting time; CanOps= no. of cancelled operations; CAN1M = I month cancer waits; CAN2M = 2 month cancer waits; THROM = time to thrombolysis; REVAS = time for revascularization; CHEST = access to rapid access chest pain clinic; Del.Trans = delayed transfers into a care setting; Choice = Government's patient choice program

6.2.2 Multivariate analysis

As a precursor to multivariate analysis, further insights into the relationships between *NFP*, *REF* and financial performance are provided by the Pearson correlation coefficients reported in Table 11. Financial performance has been measured in three ways, all scaled by total revenue: first, reported financial performance (*FP*); second, operating surplus (*OS*) which being unaffected by transactions such as asset sales and the public dividend is less noisy than reported financial performance and thirdly, underlying financial performance defined as reported financial performance less disclosed financial support (*UFP*). *UFP* is a measure of underlying financial performance is necessarily crude in that financial support was not disclosed in 2002-03 and the reporting in other years, until its cessation in 2006-07, was asymmetrical: those Trusts releasing funds for brokerage and Trusts repaying financial support were not required to make equivalent disclosures. The result is that the underlying surpluses of some Trusts in the years up to 2006-07 may be higher than reported. Nonetheless the adjustment for financial support does reflect underlying deficits more accurately. Panel A shows the results for the full sample (854 observations), Panel B shows the results for general Trusts (769 observations) and Panel C shows the results for the sample excluding Trusts which converted to Foundation Trust status during the study period (625 observations).

The association between nonfinancial performance *NFP* and the other variables is first considered. All panels of Table 11 show that there is a negative relationship between *NFP* and *REF* (0.227, Panel A). Trusts with lower costs (*REF)* and therefore higher cost efficiency have higher *NFP,* with *NFP* falling by 1 point for a 4-5 point increase in reference costs. There are also significant results between *NFP* and all three measures of financial performance. However, the effects are small (coefficient range 0.14-0.24). A deterioration in financial performance of over 4% of revenue (approximately £8m for a Trust with mean revenue of £200m) would be required to lose one *NFP* point. To put this in perspective the mean deficit for the sample is 0.9% and the standard deviation is 3.5%. This result reflects the wide variation in financial performance associated with the clustering of *NFP* at high levels, particularly in the Star ratings (see Table 7). Finally, there is a negative association between Trust size (*SIZE*) and nonfinancial performance (*NFP)*

indicating that larger Trusts, consistent with a more complex case mix, generate lower *NFP*.

For the sample of general Trusts (769 observations) there is also a negative association between Trust reference costs and all three measures of financial performance, but again the coefficients are small (c. -0.09) and the significance is lost if the full sample (854 observations) or the sample which excludes Trusts converting to Foundation Trust status (625 observations) are used. For an average Trust with £200m revenue a deficit of over £20m would be required to predict an increase from 100 to 101 in reference costs. To put this into perspective, the mean deficit of Trusts with mean revenue of £200m is approximately £2m with a standard deviation of £7m.

Table 11: Pearson correlation co-efficients

Panel A: Full sample

	NFP	*REF*	*FP*	*OS*	*UFP*	*SIZE*
NFP	1					
REF	-0.2268*	1				
FP	0.2128*	-0.0076	1			
OS	0.2271*	-0.0215	0.9320*	1		
UFP	0.1913*	-0.0060	0.8927*	0.8438*	1	
SIZE	-0.1561*	-0.0284	0.0141	0.0323	0.0035	1
No. obs .	854					

Panel B: General Trusts

	NFP	*REF*	*FP*	*OS*	*UFP*	*SIZE*
NFP	1					
REF	-0.2059*	1				
FP	0.2369*	-0.0865*	1			
OS	0.2513*	-0.0838*	0.9334*	1		
UFP	0.2211*	-0.0952*	0.8973*	0.7206*	1	
SIZE	-0.1685*	0.2794*	0.0584	0.3241*	0.0517	1
No. obs .	769					

Panel C: Full sample excluding Trusts converting to Foundation Trust status

	NFP	*REF*	*FP*	*OS*	*UFP*	*SIZE*
NFP	1					
REF	-0.2368*	1				
FP	0.1812*	-0.0002	1			
OS	0.1991*	-0.0034	0.9457*	1		
UFP	0.1409*	0.0068	0.8910*	0.8531*	1	
SIZE	-0.1145*	0.0322	0.0576	0.0718	0.0615	1
No. obs .	625					

** Significant at 5%*

Key: *NFP* is a measure of nonfinancial performance based on performance against key Government measures; *REF* is the Trust reference cost index, an inverse measure of Trust cost efficiency; *FP* is reported financial performance scaled by total revenue; *OS* is operating surplus/deficit scaled by total revenue; *UFP* is underlying financial performance defined as reported financial performance less financial support, scaled by total revenue; *SIZE* is the natural log of Trust total assets.

Specialist Trusts are identified by reference to the Healthcare Commission's classification of acute Trusts. Trusts converting to Foundation Trust status are identified by reference to Monitor, the Foundation Trust regulator and the Laing and Buisson database of NHS financial statements.

To identify the relationship between performance against Government targets and cost efficiency in each PMS *NFP* is regressed on Trust reference costs (*REF*), reported financial performance scaled by total revenue (*FP*), and size measured by the log of total assets (*SIZE*) for the whole study period (1) and then for each of the Star ratings (2) and the AHC (3).

$$NFP_{it} = \alpha_0 + \alpha_1 REF_{it} + \alpha_2 FP_{it} + \alpha_3 SIZE_{it} + \sum_{l=t}^{l=n} \alpha_l YEAR_i + e_{it} \qquad (11)$$

Where: for (1), covering the whole study period, t=2003 and n =2008; for (2), covering the period of the Star ratings, t=2003 and n=2005; and for (3), covering the period of the AHC, t= 2006 and n =2008.

The results for the whole sample are shown in Table 12 Panel A and for the sample of general Trusts in Table 12 Panel B. For each variable the point estimate, the robust standard error (in parentheses) and, italicized, the p value are reported.

Panel A shows that over the whole period there is a significant negative relationship between *NFP* and Trust reference costs (coefficient -0.069, p= 0.000) indicating that increased cost efficiency is associated with better performance against Government targets whichever PMS is in place. This result is consistent with Deily and McKay (2006) and Ludwig, Van Merode and Groot (2010) and indicates that the pursuit of higher standards of nonfinancial performance is associated with improved cost efficiency. However there are notable differences in the Star ratings (2) (coefficient -0.050, p=0.000) and the AHC (3) (coefficient -0.089, p=0.000). These results provide evidence that in the AHC, a reduction in cost efficiency is associated with a much higher loss of NFP than in the Star ratings. In the AHC an increase in *REF* of 10 points is predicted to lose almost 1 point in NFP as compared with 0.5 in the Star ratings.

The relationship between nonfinancial performance (*NFP)* and financial performance also shows differences between the two PMS. The coefficient on *FP* for the whole period is 0.231 (p=0.000). In the AHC however the coefficient is lower at 0.152 (p=0.000) and in the Stars it is higher at 0.394 (p=0.000). These results are consistent with the greater ease with which measured performance can be manipulated in the Star ratings. Panel B shows the results for the sample which excludes specialist Trusts. There are no notable differences in the results. The main

coefficient of interest is that on cost efficiency (*REF*) which is -0.067 (comparator -0.069) for the full sample, -0.047 (-0.050) for the Star ratings and -0.088 (-0.089) for the AHC. All results significant at the 1% level.

A robustness test of the results from Table 12 is provided in Table 13 which substitutes operating surplus *(OS)* for reported financial performance *(FP).* Once more there are no notable differences in the results. The coefficient on *REF* for the full sample remains at -0.069 for the full period, -0.049 for the Star ratings and -0.090 for the AHC. For the sample which excludes specialist Trusts the coefficients are only slightly different at -0.065, -0.043 and -0.088. Again all results are significant at the 1% level of significance.

These results indicate a strong and positive relationship between service standards and cost efficiency during the period of study which is robust to sample construction and choice of financial measure. The result for the Star ratings is lower than that of the AHC and is consistent with greater opportunities in the Star ratings for delivering higher levels of nonfinancial performance other than by improving cost efficiency.

Table 12: Relationship between nonfinancial performance and cost efficiency – individual regressions

Panel A: Full sample (854 observations)

	(1) Full period	(2) Star ratings	(3) AHC
VARIABLES	*NFP*	*NFP*	*NFP*
REF	-0.069***	-0.050***	-0.089***
	(0.009)	(0.012)	(0.015)
	0.000	0.000	0.000
FP	0.231***	0.394***	0.152***
	(0.033)	(0.054)	(0.044)
	0.000	0.000	0.001
SIZE	-0.490***	-0.436**	-0.523*
	(0.174)	(0.204)	(0.309)
	0.005	0.033	0.091
Year effects	YES	YES	YES
Constant	47.02***	44.54***	47.16***
	(2.299)	(2.759)	(4.026)
	0.000	0.000	0.000
Observations	854	497	357
R-squared	0.235	0.139	0.132

Panel B: General Trusts (769 observations)

	(1) Full period	(2) Star ratings	(3) AHC
VARIABLES	*NFP*	*NFP*	*NFP*
REF	-0.067***	-0.047***	-0.088***
	(0.014)	(0.018)	(0.023)
	0.000	0.009	0.000
FP	0.219***	0.334***	0.163***
	(0.032)	(0.052)	(0.042)
	0.000	0.000	0.000
SIZE	-0.376*	-0.239	-0.564*
	(0.204)	(0.249)	(0.340)
	0.065	0.338	0.098
Year effects	YES	YES	YES
Constant	45.22***	41.69***	47.49***
	(2.410)	(2.985)	(4.022)
	0.000	0.000	0.000
Observations	769	443	326
R-squared	0.231	0.115	0.129

Robust standard errors, adjusted for clustering, in parentheses; p values in italics

*** p<0.01, ** p<0.05, * p<0.1

NFP is a measure of nonfinancial performance based on performance against key Government measures; *REF* is the Trust reference cost index, an inverse measure of Trust cost efficiency; *FP* is reported financial performance scaled by total revenue; *SIZE* is the natural log of Trust total assets.

Full period: 2003-2008; Star ratings: 2003-2005; AHC: Annual Health Check 2006-2008; Specialist Trusts as defined by Healthcare Commission; Converting Trusts are identified by reference to Monitor, the Foundation Trust regulator and the Laing and Buisson database of NHS financial statements.

Table 13: Relationship between nonfinancial performance, cost efficiency and operating surplus

Panel A: Full sample

	(1) Full period	(2) Star ratings	(3) AHC
VARIABLES	*NFP*	*NFP*	*NFP*
REF	-0.069***	-0.049***	-0.090***
	(0.010)	(0.013)	(0.015)
	0.000	0.000	0.000
OS	0.231***	0.342***	0.170***
	(0.034)	(0.052)	(0.046)
	0.000	0.000	0.000
SIZE	-0.561***	-0.556***	-0.566*
	(0.175)	(0.206)	(0.309)
	0.001	0.007	0.068
Year effects	YES	YES	YES
Constant	46.91***	44.41***	47.37***
	(2.300)	(2.785)	(4.018)
	0.000	0.000	0.000
Observations	854	497	357
R-squared	0.234	0.123	0.137

Panel B: Sample excluding specialist Trusts

	(1) Full period	(2) Star ratings	(3) AHC
VARIABLES	*NFP*	*NFP*	*NFP*
REF	-0.065***	-0.0433**	-0.0885***
	(0.014)	(0.0182)	(0.0230)
	0.000	0.018	0.000
OS	0.224***	0.296***	0.184***
	(0.0327)	(0.0507)	(0.0441)
	0.000	0.000	0.000
SIZE	-0.461**	-0.365	-0.620*
	(0.205)	(0.253)	(0.340)
	0.025	0.150	0.069
Year effects	YES	YES	YES
Constant	45.22***	41.69***	47.84***
	(2.409)	(3.007)	(4.011)
	0.000	0.000	0.000
Observations	769	443	326
R-squared	0.231	0.102	0.136

Robust standard errors in parentheses; p values in italics
*** p<0.01, ** p<0.05, * p<0.1

NFP is a measure of nonfinancial performance based on performance against key Government measures; *REF* is the Trust reference cost index, an inverse measure of Trust cost efficiency; *FP* is reported financial performance scaled by total revenue; *SIZE* is the natural log of Trust total assets.

Full period: 2003-2008; Star ratings: 2003-2005; AHC: Annual Health Check 2006-2008; Specialist Trusts as defined by Healthcare Commission; Converting identified by reference to Monitor, the Foundation Trust regulator and the Laing and Buisson database of NHS financial statements.

Tables 12 and 13 show that there is a difference between the Stars and the AHC in their effectiveness to incentivise cost efficiency to improve service standards. To identify whether these differences are significant an indicator variable, PMS, coded 0 for the AHC and 1 for the Star ratings, is used to introduce interaction effects for reference costs (*PMS*REF*) and financial performance (*PMS*FP*) (see equation 12) to measure the differential impact of the Star Ratings as compared with the AHC.

$$NFP_{it} = \alpha_0 + \alpha_1 REF_{it} + \alpha_2 PMS * REF_{it} + \alpha_3 FP_{it} + \alpha_4 PMS * FP_{it} + \alpha_5 SIZE_{it} + \sum_{l=2003}^{2007} \alpha_l YEAR_l + e_{it} \qquad (12)$$

Table 14 shows the results for three samples. Panel A shows the results for the full sample, Panel B excludes specialist Trusts and Panel C additionally excludes Trusts which convert to Foundation Trust status during the study period. In each panel, Model 1 represents the main model with reported financial performance (*FP*) as the measure of financial performance. Models 2 and 3 offer robustness checks for Model 1 by substituting operating surplus (*OS*, Model 2) and underlying financial performance (*UFP*, Model 3) for reported financial performance. Underlying financial performance is measured as reported financial performance minus financial support disclosed in Trust financial statements. All measures of financial performance are scaled by total revenue. For each variable the point estimate is first reported; secondly, the robust standard error (adjusted for heteroskedasticity and clustering), and thirdly, italicized, the p value.

In Panel A, the full sample, the coefficient on *REF* for the principal model, Model 1, is -0.09. This is the coefficient for the PMS which is coded 0, and so represents the relationship between *NFP* and *REF* for the AHC. The coefficient is consistent with the results for the individual regression where the coefficient was also found to be -0.09 (See Tables 12 and 13, Panels A and B) and is robust to choice of financial measure (Models 2 and 3). The coefficient of 0.04 on *PMS*REF* indicates the differential impact of the Star ratings as compared with the AHC. Summing the coefficients on REF (the coefficient on the AHC) and on the interaction, PMS*REF, (the differential impact of the Star ratings) gives the coefficient on the Star ratings. From Table 14 this can be computed as being equal to - 0.09+0.04=-0.05. This coefficient is consistent with the results for the individual regression for the Star ratings where the coefficient was -0.05 (See Tables 12 and

13, Panels A and B). However the interaction is not significant for any of Models 1, 2 and 3 indicating that the relationship between service standards and Trust cost efficiency was not significantly different in the AHC than in the Stars. For Model 1 there is a significantly positive relationship between *NFP* and *FP* (coefficient in the AHC = 0.152, p=0.005) and the difference between the Stars and the AHC is also significant (coefficient = 0.242, p =0.022). This result indicates the relationship between NFP and FP is significantly stronger in the Stars as compared with the AHC. This result is robust to alternative measures of financial performance (Models 2 and 3).

However, when specialist Trusts are excluded, (Panel B) a different picture emerges. The coefficient on *REF* remains at -0.09 but the interaction effect *(PMS*REF)* is now slightly bigger at 0.05 and is significant at about 10% (Model 1, p=0.104; Model 2, p= 0.083; Model 3, p= 0.093). These results show that the relationship between *NFP* and *REF* in the AHC is significantly different from that in the Star ratings and is robust to choice of financial performance measure.

The relationship between nonfinancial and financial performance in Table 14 Panel B is more volatile. In model 1 the coefficient on *FP* is 0.16 and the interaction coefficient, which is significant at the 10% level, is 0.17, consistent with a coefficient of 0.33 in the Star ratings. However both the size of the coefficient and the significance of these results is sensitive to choice of financial performance measure.

Table 14 Panel C shows the results of the regressions for a sample which excludes both specialist Trusts and those which converted to Foundation Trust status during the period (578 observations). The coefficients on *REF* for Model 1 is, as in previous panels, -0.09. This is significant at the 1% level of significance and is robust to choice of financial performance measure (Models 2 and 3). The coefficient on the interaction is higher at 0.06 than in Panels B (0.05) and A (0.04). These results are indicative that the relative effect of the AHC is increased for the poorer performing Trusts in the tail of the performance comet. As in Panel B (all general Trusts) the result is significant at about 10% and is robust to choice of financial performance measure.

In summary Table 14 shows that there was a significantly positive relationship between service standards and cost efficiency in NHS acute hospital

Trusts over the period 2003-2008. Further, for general Trusts, this relationship was significantly stronger for the AHC than for the Star ratings and was stronger still for the weaker Trusts which did not convert to Foundation Trust status during the study period.

Table 14: The relationship between nonfinancial performance and Trust cost efficiency - the Star ratings as compared with the AHC

Panel A: Full sample

Dep. Variable *NFP* VARIABLES	Model 1 *(FP)*	Model 2 *(OS)*	Model 3 *(UFP)*
REF	-0.089*** (0.019) *0.000*	-0.090*** (0.019) *0.000*	-0.089*** (0.019) *0.000*
*PMS*REF*	0.038 (0.032) *0.240*	0.041 (0.033) *0.207*	0.037 (0.033) *0.264*
FP	0.152*** (0.053) *0.005*		
*PMS*FP*	0.242** (0.105) *0.022*		
OS		0.170*** (0.052) *0.001*	
*PMS*OS*		0.172* (0.096) *0.073*	
UFP			0.122*** (0.045) *0.006*
*PMS*UFP*			0.200** (0.079) *0.011*
SIZE	-0.468*** (0.172) *0.007*	-0.560*** (0.173) *0.001*	-0.437** (0.172) *0.011*
PMS	-0.536 (3.133) *0.864*	-2.027 (3.200) *0.527*	-0.134 (3.186) *0.967*
Year effects	YES	YES	YES
Constant	45.70*** (2.550) *0.000*	46.48*** (2.553) *0.000*	45.30*** (2.558) *0.000*
Observations	854	854	854
R-squared	0.248	0.243	0.245

Robust standard errors, adjusted for clustering, in parentheses; p values in italics

*** p<0.01, ** p<0.05, * p<0.1

NFP is a measure of nonfinancial performance based on performance against key Government measures; *REF* is the Trust reference cost index, an inverse measure of Trust cost efficiency ; *FP* is reported financial performance scaled by total revenue; *OS* is operating surplus/deficit scaled by total revenue; *UFP* is underlying financial performance defined as reported financial performance less financial support, scaled by total revenue; *PMS* is an indicator variable coded 0 for the AHC (2006-2008) and 1 for the Star ratings. *PMS*REF* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and cost efficiency; *PMS*FP* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and reported financial performance*: PMS*OS* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and operating surplus/deficit*: PMS*UFP* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and underlying financial performance.

Table 14: The relationship between nonfinancial performance and Trust cost efficiency - the Star ratings as compared with the AHC

Panel B: General Trusts

VARIABLES	Model 1 *NFP*	Model 2 *NFP*	Model 3 *NFP*
REF	-0.092***	-0.091***	-0.092***
	(0.023)	(0.023)	(0.023)
	0.000	*0.000*	*0.000*
*PMS*REF*	0.048	0.051*	0.049*
	(0.029)	(0.029)	(0.029)
	0.104	*0.083*	*0.093*
FP	0.160***		
	(0.053)		
	0.003		
*PMS*FP*	0.176*		
	(0.102)		
	0.084		
OS		0.182***	
		(0.052)	
		0.001	
*PMS*OS*		0.118	
		(0.095)	
		0.218	
UFP			0.136***
			(0.045)
			0.002
*PMS*UFP*			0.161**
			(0.079)
			0.041
SIZE	-0.374*	-0.472**	-0.361*
	(0.196)	(0.202)	(0.195)
	0.058	*0.020*	*0.064*
PMS	-1.802	-2.855	-2.001
	(2.867)	(2.872)	(2.880)
	0.530	*0.320*	*0.487*
Year effects	YES	YES	YES
Constant	44.82***	45.57***	44.65***
	(3.194)	(3.217)	(3.207)
	0.000	*0.000*	*0.000*
Observations	769	769	769
R-squared	0.240	0.237	0.242

Robust standard errors, adjusted for clustering, in parentheses; p values in italics
*** p<0.01, ** p<0.05, * p<0.1

NFP is a measure of nonfinancial performance based on performance against key Government measures; *REF* is the Trust reference cost index, an inverse measure of Trust cost efficiency ; *FP* is reported financial performance scaled by total revenue; *OS* is operating surplus/deficit scaled by total revenue; *UFP* is underlying financial performance defined as reported financial performance less financial support, scaled by total revenue; *PMS* is an indicator variable coded 0 for the AHC (2006-2008) and 1 for the Star ratings. *PMS*REF* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and cost efficiency; *PMS*FP* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and reported financial performance*: PMS*OS* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and operating surplus/deficit*: PMS*UFP* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and underlying financial performance.

Table 14: The relationship between nonfinancial performance and Trust cost efficiency - the Star ratings as compared with the AHC

Panel C: General Trusts excluding those converting to Foundation status

VARIABLES	Model 1 *NFP*	Model 2 *NFP*	Model 3 *NFP*
REF	-0.0908*** (0.0255) *0.000*	-0.0903*** (0.0256) *0.000*	-0.0909*** (0.0256) *0.000*
*PMS*REF*	0.0556 (0.0346) *0.109*	0.0573* (0.0344) *0.096*	0.0546 (0.0348) *0.117*
FP	0.137** (0.0558) *0.014*		
*PMS*FP*	0.145 (0.108) *0.179*		
OS		0.160*** (0.0553) *0.004*	
*PMS*OS*		0.106 (0.101) *0.292*	
UFP			0.111** (0.0485) *0.022*
*PMS*UFP*			0.127 (0.0884) *0.152*
SIZE	-0.461* (0.239) *0.054*	-0.559** (0.245) *0.023*	-0.442* (0.237) *0.063*
Year effects	YES	YES	YES
Constant	45.79*** (3.652) *0.000*	46.56*** (3.683) *0.000*	45.59*** (3.659) *0.000*
Observations	578	578	578
R-squared	0.187	0.189	0.182

Robust standard errors, adjusted for clustering, in parentheses; p values in italics
*** p<0.01, ** p<0.05, * p<0.1

NFP is a measure of nonfinancial performance based on performance against key Government measures; *REF* is the Trust reference cost index, an inverse measure of Trust cost efficiency ; *FP* is reported financial performance scaled by total revenue; *OS* is operating surplus/deficit scaled by total revenue; *UFP* is underlying financial performance defined as reported financial performance less financial support, scaled by total revenue; *PMS* is an indicator variable coded 0 for the AHC (2006-2008) and 1 for the Star ratings. *PMS*REF* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and cost efficiency; *PMS*FP* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and reported financial performance*: PMS*OS* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and operating surplus/deficit*: PMS*UFP* is an interaction effect to identify the differential impact of the Star ratings as compared with the AHC on the relationship between Trust nonfinancial performance and underlying financial performance.

6.2.3 Summary of findings: Study 1 – the impact of PMS design on the incentivisation of cost efficiency

The aim of this study is to investigate the impact of PMS design on the incentivisation of cost efficiency as a means of improving nonfinancial performance. The research focus is a comparison of the relative impact of the AHC as compared with the Star ratings, applied to NHS hospital Trusts in the period 2003-2008. The AHC was characterised by better balance in the weightings given to financial and nonfinancial measures and by greater measurement diversity. A review of these characteristics, facilitated by the production possibility frontier analysis of Chapter 5, gave rise to the prediction that in the AHC the incentive to trade financial for nonfinancial performance would be reduced and that the AHC would be more effective at incentivising cost efficiency than the Star ratings. Categorical and non-parametric chi–square analysis provides evidence of manipulation consistent with the manipulation of financial performance in order to improve overall performance ratings and supports the prediction that the prevalence of deficits would be lower in the AHC than in the Stars. The reduction in deficits was most in evidence for small deficits of up to 1% of revenue, which reduced from 12% to 2% of all observations. A reduction in large deficits of more than 1% of revenue however was not in evidence. These remained at a constant 30% of all observations. These findings are consistent with the characteristics of the AHC which heavily penalised deficits but once a deficit was incurred there was no further penalty. Thus, unless a Trust could move into surplus, which practically would be more difficult the larger the deficit, there was no incentive to take action.

Evidence from multivariate analysis provides evidence that the AHC was more effective at incentivising cost efficiency than the Stars, consistent with it being a more robust performance measurement system which offers less scope for manipulation and which incorporates stronger incentives to improve cost efficiency as a means of improving service standards. Using separate regressions for each of the Stars and the AHC the coefficient on *REF,* a measure of Trust cost efficiency, was higher in the AHC than in the Stars and both results were statistically significant and robust to sample selection and choice of financial measure. When interaction effects were introduced to identify the significance of the differences between the AHC and the Stars, the coefficients remained consistent with the individual regressions and for the sample of general hospitals the difference was identified as

being significant. These results were robust to alternative measures for financial performance and to the exclusion of Trusts converting to Foundation Trust status. Indeed the size of the coefficient on REF for the sample excluding converting Trusts was higher, indicating that the relative impact of the AHC on the incentivisation of cost efficiency was bigger for weaker Trusts.

It has not been possible to explore the impact of the AHC and the Stars on specialist Trusts alone as there are too few observations. However, descriptive statistics shows that they are significantly smaller and less efficient[50] than general (non-specialist) Trusts but have better reported financial performance.

[50] The observed lower efficiency of specialist Trusts is consistent with the findings of prior research using frontier analysis techniques (Worthington, 2004).

6.3 FINDINGS STUDY 2: AN INVESTIGATION OF THE PERFORMANCE IMPACT OF FINANCIAL SUPPORT

6.3.1 Descriptive statistics

Financial support was disclosed in financial statements over the three year period from 2003-04 to 2005-06. The total number of Trust observations in these years is 461 (see Table 1). Eight observations obtained from the Laing and Buisson database are eliminated from this population: five observations where an underlying surplus was associated with the receipt of financial support and three observations of negative financial support. The resulting core sample for this study is therefore 453 observations. The results from this core sample are subject to additional tests to identify their robustness to the inclusion of these eight omitted observations.

The impact of financial support on the numbers of Trusts reporting surpluses and deficits is revealed in Table 15. Panel A shows that, of the 280 Trusts reporting a surplus, 97 (35%) had an underlying deficit. This figure of 35% provides an indication of the impact of financial support on key performance measures, such as the fulfilment of the statutory duty to breakeven and Trust performance ratings, which are used for the purposes of public accountability. Table 15 Panel B shows that of the 162 Trusts that received financial support 97 (60% of 162) were moved into surplus whilst the remainder (65) remained in deficit. Panel C identifies Trusts not in receipt of financial support: 108 deficit Trusts, that provide the reference group for measuring benchmark nonfinancial performance, and 183 Trusts reporting surpluses. Panels B and C provide the framework for the investigation of the performance impact of financial support.

Table 15: Distribution of reported and underlying financial performance, 2003-04 to 2005-06: an analysis of the impact of financial support

Panel A: Full sample of Trusts

	Underlying financial performance *(UFP)*		
Reported financial performance *(FP)*	**Deficit N**	**Surplus N**	**Total N**
Deficit	173	0	173
Surplus	97	183	280
Total	270	183	453

Panel B: Trusts receiving financial support

	Underlying financial performance *(UFP)*		
Reported financial performance (*FP*)	**Deficit N**	**Surplus N**	**Total N**
Deficit	65	0	65
Surplus	97	0	97
Total	162	0	162

Panel C: Trusts not receiving financial support

	Underlying financial performance *(UFP)*		
Reported financial performance (*FP*)	**Deficit N**	**Surplus N**	**Total N**
Deficit	108	0	108
Surplus	0	183	183
Total	108	183	291

FP is reported financial performance scaled by total revenue; *UFP* is underlying financial performance defined as reported financial performance less financial support, scaled by total revenue. Source: Laing and Buisson

Reported financial performance is surplus/deficit reported in NHS Trust income statements and is a measure of residual income after a capital charge based on current value of assets.
Financial support is defined as funds transferred into Trusts in financial difficulty from surplus funds elsewhere in the NHS. It was not part of the revenue earned by the Trust under its service level agreements.

Table 16 illustrates the economic significance of the 453 Trusts in the sample. Over the three year period the mean revenue of each Trust amounted to £185m. However the mean revenue of specialist Trusts (49) was much lower at £82m. The mean revenue of the 162 Trusts (36% of 453) receiving financial support is somewhat larger at over £203m, than the full sample. In contrast, the revenue of specialist Trusts (4) receiving financial support is lower at £69m than the size of all (49) specialist Trusts (£82m). The difference in means (not tabulated) between the revenue of Trusts receiving financial support and not receiving financial support is significant at 5% for the whole sample (453) but not for the subsamples of acute trusts (404) and specialist trusts (49). These differences between specialist and non-specialist Trusts warrants the investigation not only of the total sample of 453 observations but also, as for Study 1, a sample excluding specialist Trusts (404 observations). For Trusts receiving financial support (162) there is no significant difference between the mean revenue of those which subsequently report a deficit and those which report a surplus. This is also the case when specialist Trusts are excluded from the sample (158 observations). However, of the four specialist Trusts receiving financial support, none subsequently reported a deficit. All four were moved from deficit into surplus.

Table 16: Trust total revenue £000, 2003-04 to 2005-06: a comparison of Trusts receiving financial support and those not receiving it.

Panel A: All Trusts

Year	N	Mean	Std.Devn.	Min.	Max.	Median
2004	172	170,975	104,838	11,523	627,148	138,895
2005	146	185,007	109,035	12,981	677,981	154,001
2006	135	201,832	120,486	42,855	721,415	164,237
Total	453	184,693	111,512	11,523	721,415	153,124
General	404	197,129	110,530	56,286	721,415	162,715
Specialist	49	82,158	49,917	11,523	220,075	64,487

Panel B: Trusts in Receipt of Financial Support

Year	N	Mean	Std.Devn.	Min.	Max.	Median
2004	67	189,901	104,710	75,142	627,148	153,758
2005	56	213,640	119,954	50,840	677,981	180,010
2006	39	211,413	121,215	57,263	558,478	189,152
Total	162	203,286	114,043	50,840	677,981	171,208
General	158	206,677	113,397	68,162	677,981	173,606
Specialist	4	69,343	22,428	50,840	101,265	62,634
Surplus	97	204,854	122,629	50,840	677,981	164,673
Deficit	65	200,947	100,756	68,162	627,148	177,378

Panel C: Trusts Not in Receipt of Financial Support

Year	N	Mean	Std.Devn.	Min.	Max.	Median
2004	105	158,898	103,613	11,523	538,363	129,640
2005	90	167,190	98,184	12,981	513,404	146,009
2006	96	197,939	120,608	42,855	721,415	160,310
Total	291	174,342	108,905	11,523	721,415	145,712
General	246	190,997	108,437	56,286	721,415	155,882
Specialist	45	83,297	51,650	11,523	220,075	64,487
Surplus	183	177,163	116,175	11,523	721,415	145,868
Deficit	108	169,562	95,661	41,968	474,983	142,020

Financial support represented funds transferred into Trusts in financial difficulty from surplus funds elsewhere in the NHS. It was not part of the revenue earned by the Trust under its service level agreements. Source: Laing and Buisson

Descriptive statistics relating to the reported financial performance of Trusts over the Study 2 period of 2003-04 to 2005-06 are reported in Table 17. The results for all Trusts are shown in Panel A and the sample is then subdivided into Trusts receiving financial support (Panel B) and Trusts not receiving financial support (Panel C). The mean deficit for all Trusts (453) is £2.3m (Panel A). The annual mean deficit increased from under £1m in 2004 to £4.5m in 2006. However, consistent with the discipline of the financial breakeven requirement, median financial performance was marginally positive in all years. Trusts receiving financial support (162, Panel B) reported a mean deficit of £2.4m and those not receiving support, (291, Panel C), reported a mean deficit of £2.3m. The null hypothesis of no difference in the means of reported performance for Trusts not in receipt of financial support compared to those receiving financial support (untabulated) is not rejected ($t = 0.0581$; $p = 0.479$), indicating the effectiveness of financial support in narrowing the difference (£32k) in reported financial performance between these two groups. Panels B and C also show the results for Trusts reporting deficits and those reporting surpluses. As for the full sample there is no significant difference (untabulated) in the mean reported financial performance for Trusts reporting deficits and for Trusts reporting surpluses whether or not they receive financial support. In Panel D, as a control for heteroskedasticity in the multivariate analysis, financial performance is scaled by total revenue. Consistent with Panel A, the mean reported financial performance is a deficit of 1.5% of revenue, with a small positive median of 0.003% of revenue and a range from a maximum deficit of 28% to a maximum surplus of 6.5%. There is no significant difference between those receiving financial support and those not.

Table 17: Reported financial performance, 2003-04 to 2005-06: a comparison of Trusts receiving financial support and those not receiving it

Panel A: All Trusts

Year	N	Mean	Std. Devn.	Minm.	Maxm.	Median
		£000	£000	£000	£000	£000
2004	172	-826	2,583	- 18,637	9,394	5
2005	146	-2,112	4,946	- 30,657	13,581	3
2006	135	-4,531	7,952	- 40,281	6,481	2
Total	**453**	**-2,345**	**5,607**	**- 40,281**	**13,581**	**4**

Panel B: Trusts receiving financial support

Year	N	Mean	Std. Devn.	Minm.	Maxm.	Median
		£000	£000	£000	£000	£000
2004	67	- 1,332	2,911	- 18,637	513	3
2005	56	- 2,294	5,743	- 30,657	13,581	7
2006	39	- 4,241	8,353	- 40,281	6,481	12
Total	**162**	**- 2,365**	**5,701**	**- 40,281**	**13,581**	**4**
Surplus	97	450	1,613	-	13,581	50
Deficit	65	-6,566	6,924	- 40,281	- 27	-4506

Panel C: Trusts not receiving financial support

Year	N	Mean	Std.Devn.	Minm.	Maxm.	Median
		£000	£000	£000	£000	£000
2004	105	-502	2,306	-12,801	9,394	6.000
2005	90	-1,999	4,410	-21,656	10,755	1.500
2006	96	-4,649	7,825	-33,569	3,414	0.500
Total	**291**	**-2,333**	**5,564**	**-33,569**	**10,755**	**4.000**
Surplus	183	406	1,271	0	10,755	28.0
Deficit	108	-6,975	6,824	-33,569	-18	-4952.5

Panel D: Reported financial performance scaled by total revenue

	N	Mean	Std. Devn.	Minm.	Maxm.	Median
		£000	£000	£000	£000	£000
Total sample	**453**	**-1.459**	**3.487**	**-28.027**	**6.501**	**0.003**
Financial support	162	-1.393	3.587	-28.027	5.934	0.003
No financial support	291	-1.495	3.435	-22.094	6.501	0.003

Source: Laing and Buisson

Reported financial performance is surplus/deficit reported in NHS Trust income statements and is a measure of residual income after a capital charge based on current value of assets.

Financial support is defined as funds transferred into Trusts in financial difficulty from surplus funds elsewhere in the NHS. It was not part of the revenue earned by the Trust under its service level agreements.

The amounts of financial support received by Trusts and the impact it has on reported financial performance is further explored in Table 18. Panel A shows that the mean amount of financial support received by 162 Trusts (36% of the total 453) amounted to £5.7m and that this increased from £4.8m in 2004 to £6.9m in 2006. There is no significant difference (untabulated) between financial support received by Trusts reporting a surplus (£5.9m) and Trusts reporting a deficit (£5.3m). The amount of support ranged from relatively small amounts of circa £200k to a maximum of £32.2m. Panel B shows the significant impact of financial support on reported financial performance with mean financial support of £5.7m converting a mean underlying deficit of £8.0m to a mean reported deficit of £2.4m. In aggregate, over the three year disclosure period, financial support received by NHS Trusts amounted to £919m with £573m going to Trusts which subsequently reported a surplus and £346m to Trusts which subsequently remained in deficit. It is notable that the mean underlying financial deficit of Trusts reporting a surplus is, at £5.5m, half the level of Trusts which subsequently reported a deficit (£11.9m). This is in contrast to the figures for the amount of financial support received for which there is no significant difference (untabulated). It is not possible to identify from these figures whether the Trusts which subsequently reported a deficit had originally forecast breakeven, and failed to achieve it, consistent with a failure to identify good quality managers (potentially as a consequence of opportunism), or whether good quality managers were correctly selected but were given limited access to financial support which was insufficient to eliminate the underlying deficit. Further insight into this question can be provided by the multivariate analysis which follows. If financial support was provided to protect service standards then its provision will be associated with a higher level of nonfinancial performance than would otherwise be the case. This will be the case whether or not the Trust subsequently reported a deficit or a surplus.

Table 18: Financial support: amounts received by Trusts and impact on reported financial performance

Panel A: Amount of financial support received by Trusts

Year	N	Mean	Std. Devn.	Minm.	Maxm.	Median	Skewness	Kurtosis
		£000	£000	£000	£000	£000		
2004	67	4,782	5,962	250	32,210	2,950	2.752	11.161
2005	56	5,886	5,139	209	26,377	4,537	1.847	6.938
2006	39	6,894	5,037	298	19,272	5,481	0.850	2.901
Total	**162**	**5,672**	**5,504**	**209**	**32,210**	**3,985**	**2.025**	**7.925**
Surplus	97	5,908	6,199	209	32,210	3,800	2.050	7.360
Deficit	65	5,320	4,284	400	19,272	4,300	1.234	4.412

Panel B: Impact of Financial Support on Reported Financial Performance

Year	N	% Trusts	Mean reported income	Mean financial support	Underlying financial performance	Aggregate financial support
			£000	£000	£000	£000
2004	67	39%	-1,332	4,782	-6,114	320,377
2005	56	38%	-2,294	5,886	-8,180	329,615
2006	39	29%	-4,241	6,894	-11,135	268,880
Total	**162**	**36%**	**-2,365**	**5,672**	**-8,037**	**918,872**
Surplus	97	**21%**	450	5,908	**-5,458**	**573,043**
Deficit	65	**14%**	-6,566	5,320	**-11,886**	**345,829**

Reported financial performance is residual income after a capital charge based on current value of assets reported in NHS Trust income statements; Underlying financial performance is reported financial performance less financial support received; Financial support is a mechanism where additional 'revenue' was credited to a Trust in deficit from elsewhere in the NHS.

Descriptive statistics for nonfinancial performance (NFP) are reported in Table 19. The mean performance for all Trusts is 33.5 relative to a maximum score of 36. The minimum score is 5. The mean (33.2) and median (34) performance for the 162 Trusts receiving financial support is similar to the performance of those not receiving financial support (33.6 and 35 respectively). The null hypothesis of no difference in mean *NFP* score is not rejected (t=1.22; p=0.11). There is also a high degree of similarity between the performance of Trusts which reported a deficit whether or not they received financial support.

Table 19: Analysis of nonfinancial performance for Trusts in receipt of financial support as compared with those not receiving financial support

NFP	**N**	**Mean**	**Std. Devn.**	**Minm.**	**Maxm.**	**Median**	**Skewness**	**Kurtosis**
Trusts receiving financial support								
Remaining in deficit	65	32.6	3.8	21	36	33	-1.194	3.646
Converted to surplus	97	33.6	2.8	23	36	35	-1.389	4.992
	162	***33.2***	***3.2***	***21***	***36***	***34***	***-1.404***	***4.684***
Trusts not receiving financial support								
Reporting a deficit	108	32.3	4.5	5	36	34	-2.758	14.753
Reporting a surplus	183	34.4	2.5	22	36	36	-2.283	9.347
	291	***33.6***	***3.5***	***5***	***36***	***35***	***-3.125***	***19.433***
Total	***453***	***33.5***	***3.4***	***5***	***36***	***34***	***-2.594***	***15.110***

NFP is a measure of nonfinancial performance based on performance against key Government measures (Source: Healthcare Commission); Financial support is defined as funds transferred into Trusts in financial difficulty from surplus funds elsewhere in the NHS. It was not part of the revenue earned by the Trust under its service level agreements (Source: Laing and Buisson).

Based on the above analysis, Table 20 reports the results of one tailed t-tests which show whether the nonfinancial performance *(NFP)* of Trusts receiving financial support was higher than for those Trusts which did not receive financial support. As only Trusts in deficit received financial support, the comparison is made with deficit Trusts not receiving financial support, in order to ensure a like-for-like comparison. Panel A shows that across the sample of 270 observations the nonfinancial performance of Trusts receiving financial support was significantly higher at 33.2 than for those deficit Trusts not receiving financial support (t=-1.88, p=0.031). Further insight into the impact of financial support on these Trusts is obtained by subdividing the sample further into support-receiving Trusts which subsequently reported a surplus (Panel B) and those reporting a deficit (Panel C). Panel B shows that Trusts receiving support and subsequently reporting a surplus performed significantly better, at *NFP* =33.6, than those deficit Trusts which did not receive financial support, *NFP*=32.3 (t=2.39, p=0.009). This result is in contrast to the result for Trusts which remained in deficit (Panel C). For these Trusts there is no significant difference (t=-0.472, p=0.319) between those Trusts who received financial support (*NFP* =32.6) and those that did not (*NFP*=32.3).

These results provide prima facie evidence that, for the majority of Trusts in receipt of financial support, (those that reported a surplus), the receipt of financial support is associated with a significant increase in service standards. However there is a minority of Trusts for which the receipt of financial support is not associated with an improvement in nonfinancial performance. This result suggests that, for this minority of Trusts, those that remained in deficit, the selection of 'good' managers was flawed such that 'poor' managers were able to access financial support. This question is investigated further in the multivariate analysis which follows.

Table 20: One tailed t-tests comparing nonfinancial performance of Trusts receiving financial support and those not, 2003-04 to 2005-06

Panel A: Trusts reporting a deficit not receiving financial support and all Trusts receiving financial support

NFP	No. Obs.	Mean	Std. Err.	Std. Dev.
No FS	108	32.3	0.4	4.5
FS	162	33.2	0.3	3.2
Total	270	33	0.2	3.8
Difference		-0.9	0.5	
t-statistic		-1.880		
p (FS>No FS)		0.031		

Panel B: Trusts reporting a deficit not receiving financial support and Trusts receiving financial support, reporting a surplus

NFP	No. Obs.	Mean	Std. Err.	Std. Dev.
No FS	108	32.3	0.4	4.5
FS	97	33.6	0.3	2.8
Total	205	32.9	0.3	3.8
Difference		-1.3	0.5	
t-statistic		-2.391		
p (FS>No FS)		0.009		

Panel C: Trusts reporting a deficit not receiving financial support and Trusts receiving financial support, reporting a deficit

NFP	No. Obs.	Mean	Std. Err.	Std. Dev.
No FS	108	32.3	0.4	4.5
FS	65	32.6	0.5	3.8
Total	173	32.4	0.3	4.3
Difference		-0.3	0.7	
t-statistic		-0.472		
p (FS>No FS)		0.319		

NFP is a measure of nonfinancial performance based on performance against key Government measures (Source: Healthcare Commission); FS = *NFP* for Trusts in receipt of financial support; No FS = *NFP* for Trusts not in receipt of financial support.

6.3.2 Multivariate analysis

As a precursor to multivariate analysis Pearson correlation coefficients for the relationships between nonfinancial performance (*NFP)*, underlying financial performance *(UFP)* Trust cost efficiency *(REF)*, and size are reported in Table 21. Panel A reports the coefficients and their significance (in italics), for the reference group of 108 Trusts reporting a deficit but not receiving financial support. These Trusts form the basis for generating a benchmark level of nonfinancial performance (*NFP*) for Trusts in receipt of financial support. For this reference group of Trusts underlying financial performance is identical to reported financial performance. As can be seen, there is a significant and negative association between *NFP* and *REF* (p=0.000), indicating that the better a Trust's cost efficiency the better its nonfinancial performance. These results are consistent with those in Study 1.

Pearson correlation coefficients for the 162 Trusts with underlying deficits and receiving financial support are reported in Panel B. Panel B includes an additional variable, *FS*, financial support. As with Panel A, there is a negative association between nonfinancial performance (*NFP)* and Trust reference costs (*REF),* (p=0.046), Trust size (p=0.047) and underlying financial performance (p=0.001). The relationship between financial support and nonfinancial performance is however negative but insignificant. The negative association of *FS* with *UFP* (p=0.000) is consistent with the use of financial support to reduce and often eliminate underlying deficits.

The rest of this section proceeds to present the findings from the investigation of the three questions identified in Chapter 5: first, was financial support associated with an improvement in nonfinancial performance; second, consistent with on-going benefits arising from the provision of financial support, was an increase in nonfinancial performance in the subsequent period associated with the receipt of financial support, and third, consistent with the provision of financial support being conditional upon a credible recovery plan, was financial support associated with an improvement in underlying financial performance in the following period.

Table 21: Pearson correlation coefficients

Panel A: Trusts in deficit not receiving financial support

108 Trusts	*NFP*	*UFP*	*REF*	*SIZE*
NFP	1			
UFP	0.148	1		
	0.127			
REF	-0.347	0.041	1	
	0.000	*0.676*		
SIZE	-0.0976	0.069	0.0752	1
	0.3148	*0.4777*	*0.4392*	

Panel B: Trusts with underlying deficits receiving financial support

162 Trusts	*NFP*	*FS*	*UFP*	*REF*	*SIZE*
PMStot	1				
FS	-0.087	1			
	0.273				
UFP	0.271	-0.680	1		
	0.001	*0.000*			
REF	-0.157	0.019	0.075	1	
	0.046	*0.807*	*0.344*		
SIZE	-0.157	-0.200	0.126	0.122	1
	0.047	*0.011*	*0.110*	*0.124*	

p values in italics

Where: *NFP* = The key nonfinancial targets in the Star ratings and the AHC are identified in endnotes 3 and 4 respectively and the methodology applied to derive uniform measures of nonfinancial performance NFP is reported in endnote 12; FS = Financial support as a % of total revenue; FP = Reported financial performance; UFP = Underlying financial performance calculated by deducting financial support from reported financial performance; REF = Trust reference cost index, an inverse measure of Trust cost efficiency; SIZE = Log of total assets

6.3.2.1 Research question 1: Was the receipt of financial support associated with an improvement in contemporaneous nonfinancial performance?

Table 22 reports the results of the investigation into the impact of financial support on contemporaneous nonfinancial performance. A benchmark level of nonfinancial performance was first obtained by applying the following OLS regression to the 101 deficit Trusts not receiving financial support:

$$NFP_{ijt} = \alpha_0 + \alpha_1 REF_{ijt} + \alpha_2 UFP_{ijt} + \alpha_3 SIZE_{ijt} + \sum_{l=2005}^{2006} \alpha_l YEAR_i + e_{ijt}$$

Where NFP_{ijt} = nonfinancial performance for Trust i in year t for the reference group j; REF = Trust cost efficiency measured by reference cost indices; UFP = underlying financial performance, which in this reference group, is identical to reported financial performance; $SIZE$ = control variable measured as the log of total assets; $YEAR$ is an indicator variable coded 1 to 3 for the years 2004-2006.

The regression coefficients $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ and year effects were then applied to the explanatory variables of Trusts receiving financial support to obtain a performance adjusted benchmark level of nonfinancial performance, as follows:

$$\hat{NFP}_{it} = \hat{\alpha}_0 + \hat{\alpha}_1 REF_{it} + \hat{\alpha}_2 UFP_{it} + \hat{\alpha}_3 SIZE_{it} + \sum_{l=2005}^{2006} \hat{\alpha}_l YEAR_l$$

Where $\hat{NFP}$ is the performance adjusted benchmark level of nonfinancial performance. After insertion of coefficient values $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2$ *and* $\hat{\alpha}_3$ and year effects, the benchmark level of nonfinancial performance was computed as follows:

For the sample of Trusts including specialist Trusts:

$$\hat{NFP}_{it} = 49.412 - 0.139\ REF_{it} + 0.051\ UFP_{it} - 0.205 SIZE_{it} + 0.959 * 2005 - 2.118 * 2006$$

And for the sample excluding specialist Trusts:

$$\hat{NFP}_{it} = 45.148 - 0.065\ REF_{it} - 0.025\ UFP_{it} - 0.497 SIZE_{it} + 1.437 * 2005 - 2.460 * 2006$$

The difference between the fitted and actual values was then calculated:

$$resNFP = NFP - \hat{NFP}$$

Where $resNFP$ is the nonfinancial performance residual.

The mean and median values of the residuals of nonfinancial performance, $resNFP$, calculated as the difference between the actual and benchmark levels of performance, are reported in Table 22 together with their p and z values for one and two tailed tests respectively. The null hypotheses of the mean and the median being equal to zero can be rejected, p=0.003 and z=0.000. The coefficient on the mean (0.68) indicates that financial support equivalent to 1% of revenue would result in an increase of 0.68 in the NFP score. Mean financial support of £5.7m for a Trust with mean revenue of £200m would thus result in an increase of 2 points in the mean NFP score. These results are consistent with the proposition that, across the whole sample of Trusts in receipt of financial support, contemporaneous nonfinancial performance is better than it would have been without the application of financial support. This finding is consistent with the effective selection of skilled managers who use the flexibility accorded by financial support to maintain high standards of service. However, when the sample is split between those Trusts which reported a surplus and those which remained in deficit a notable difference in results is observed. For those Trusts which reported a surplus mean $resNFP$ is significantly increased to 1.144 (p=0.000) whilst, for those remaining in deficit the null hypotheses of the mean and median of $resNFP$ being equal to zero cannot be rejected (p=0.515, z=0.394). For an average Trust with £200m revenue, receiving mean financial support of £5.7m and subsequently reporting a surplus, the result would be an increase of 3.3 points in the *NFP* score whereas there would be no increase in the *NFP* score of a Trust remaining in deficit. Similar results can be found for the median $resNFP$ score. This is significantly higher than the benchmark for the full group (1.393, z=0.00), higher still for Trusts reporting a surplus (1.656, z=0.000) but lower, and not significant for Trusts reporting a deficit (0.774, z=0.394). These results are robust to the exclusion of specialist Trusts (Table 22, Panel B) and the inclusion of the eight eliminated observations (untabulated).

These findings are consistent with the results of the earlier t-tests reported in Table 20. They suggest that while, overall, the receipt of financial was associated

with improved service standards, there was a significant minority of Trusts where such benefits were not realised, either because the system of selecting Trusts for the receipt of financial support was not sufficiently robust, or because the system was being used opportunistically to disguise deficits without the requisite effort being put into maintaining service standards. However, it is possible that benefits from the receipt of financial support take time to be realised. This question is the subject of the next part of the investigation of the performance impact of financial support.

Table 22: Tests of nonfinancial performance residuals for Trusts receiving financial support

Full sample

	N	Mean	p-value*	Median	No. of obs. Median>0	z-value**
Trusts receiving financial support	**162**	**0.679**	**0.003**	**1.393**	**104**	**0.000**
Converted to surplus	97	1.144	0.000	1.656	65	0.000
Remaining in deficit	65	-0.016	0.515	0.774	39	0.394

Excluding specialist Trusts

	N	Mean	p-value*	Median	No. of obs. Median>0	z-value**
Trusts receiving financial support	**158**	**0.392**	**0.060**	**1.393**	**101**	**0.004**
Converted to surplus	93	0.827	0.003	1.656	61	0.001
Remaining in deficit	65	-0.231	0.701	0.774	40	0.528

***** One tailed Test

****** Wilcoxon Sign Rank Test

To obtain coefficients for the estimation of benchmark levels of nonfinancial performance the following OLS regression (Equation 6) applied to Trusts where reported financial performance is a precise measure of underlying financial performance, that is, the 101 Trusts not receiving financial support that reported a deficit.

$$NFP_{ijt} = \alpha_0 + \alpha_1 REF_{ijt} + \alpha_2 UFP_{ijt} + \alpha_3 SIZE_{ijt} + \sum_{l=2005}^{2006} \alpha_l YEAR_i + e_{ijt}$$

NFP_{ijt} = nonfinancial performance for Trust i in year t for the reference group j; REF = Trust cost efficiency measured by reference cost indices; UFP = underlying financial performance; $SIZE$ = control variable measured as the log of total assets; $YEAR$ is an indicator variable coded 1 to 3 for the years 2004-2006.

The regression coefficients $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ and year effects are applied to the explanatory variables of the rest of the sample to obtain a performance adjusted benchmark level of nonfinancial performance for other Trusts in our sample.

Performance adjusted benchmark levels of nonfinancial performance $\hat{NFP}$ are obtained from:

$$\hat{NFP}_{it} = \hat{\alpha}_0 + \hat{\alpha}_1 REF_{it} + \hat{\alpha}_2 UFP_{it} + \hat{\alpha}_3 SIZE_{it} + \sum_{l=2005}^{2006} \hat{\alpha}_l YEAR_l$$

Which, after insertion of coefficient values $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2$ *and* $\hat{\alpha}_3$ and year effects, for the sample which includes specialist Trusts becomes:

$$\hat{NFP}_{it} = 49.412 - 0.139\, REF_{it} + 0.051\, UFP_{it} - 0.205 SIZE_{it} + 0.959 * 2005 - 2.118 * 2006$$

And for the sample excluding specialist Trusts:

$$\hat{NFP}_{it} = 45.148 - 0.065\, REF_{it} - 0.025\, UFP_{it} - 0.497 SIZE_{it} + 1.437 * 2005 - 2.460 * 2006$$

The difference between the fitted and actual values is then calculated:

$$resNFP = NFP - \hat{NFP}$$

6.3.2.2 Research question 2: Was the receipt of financial support associated with an improvement in nonfinancial performance in the subsequent period?

The results for the second research question which investigates the impact of financial support on nonfinancial performance in the subsequent period are shown in Table 23 for the full sample of Trusts receiving financial support and in Table 24 for a sample which excludes specialist Trusts. In each Table three models are presented: Model (1) reports the results for all Trusts in the sample, Model (2) reports the results for Trusts which subsequently report a surplus, and Model (3) reports the results for Trusts which subsequently report a deficit.

Table 23, Model (1) shows that the change in nonfinancial performance (ΔNFP), for the whole sample including specialist Trusts, is positively associated with financial support at the level of 5% significance (coefficient = 0.149, p=0.043). The change in *NFP* for a Trust with revenue of £200m receiving mean financial support of £5.7m is estimated at approximately 0.4. This is over and above the contemporaneous improvement in *NFP* associated with the receipt of financial support (2 points) taking the total benefit to nonfinancial performance to 2.4. This suggests that although the main benefits of financial support were realised in the year of receipt, improvements, albeit smaller, continued to flow through into the subsequent period. These results are consistent with the view that financial support allowed Trusts time to address financial problems through process improvements, which had continuing benefits for service standards, rather than through cost cutting which might be more damaging to future capacity. Inclusion of the eight eliminated observations makes no difference to these results (untabulated).

However, as in previous analysis, there are notable differences between Trusts which subsequently reported a surplus (Model 2) and those which reported a deficit (Model 3). The results for Trusts which subsequently reported a surplus are consistent with the results for the full sample. However the size of the coefficient is slightly smaller at 0.117 as compared with 0.149 suggesting that for an average sized Trust receiving average financial support, the *NFP* score would rise by a further 0.3 points as compared with 0.4. There is no association between financial support and subsequent *NFP* for Trusts reporting a deficit. Combined with the finding of no contemporaneous benefits this result suggests that no benefits to

service standards were realised in these Trusts as a consequence of receiving financial support.

Table 24 shows the results for a sample which excludes specialist Trusts. As can be seen the results are virtually identical. The coefficient on financial support is 0.149 (0.149) for the full sample, 0.124 (0.117) for surplus Trusts and 0.284 (0.284) for deficit Trusts. As previously, the results for the full sample and those reporting a surplus were significant (p=0.067 and p=0.074 respectively) whilst those for Trusts reporting a deficit were not (p=0.268). Inclusion of the eight eliminated observations (untabulated) makes no difference to these results.

In summary, these findings are consistent with the results of the earlier tests. They suggest that while, overall, the receipt of financial support was associated with improved service standards, the benefits were concentrated in those Trusts that reported a surplus. A significant minority of Trusts which reported a deficit demonstrated no improvement in service standards as a consequence of receiving financial support suggesting that the system of selecting Trusts for the receipt of financial support was not sufficiently robust, or that the system was being used opportunistically to disguise deficits without the requisite effort being put into maintaining service standards.

Table 23: Association between financial support and subsequent changes in nonfinancial performance

	Full sample	Reporting surplus	Reporting deficit
	(Model 1)	(Model 2)	(Model 3)
VARIABLES	ΔNFP	ΔNFP	ΔNFP
FS	0.149**	0.117**	0.284
	(0.073)	(0.060)	(0.253)
	0.043	*0.057*	*0.268*
ΔREF	-0.112**	-0.076*	-0.187*
	(0.045)	(0.043)	(0.099)
	0.015	*0.081*	*0.066*
ΔUFP	-0.054	-0.052	-0.109
	(0.062)	(0.062)	(0.099)
	0.387	*0.408*	*0.278*
ΔSIZE	-5.636	-1.621	-10.459
	(4.000)	(3.252)	(6.267)
	0.162	*0.620*	*0.103*
2005	-0.859	-1.396	-0.292
	(0.769)	(0.846)	(1.607)
	0.267	*0.104*	*0.857*
2006	-3.458***	-3.237***	-4.137
	(0.900)	(1.019)	(1.793)
	0.000	*0.002*	*0.026*
2007	omitted	omitted	omitted
Constant	0.646	0.669	0.728
	(0.794)	(0.831)	(1.550)
	0.418	*0.424*	*0.641*
Observations	156	92	64
R-squared	0.191	0.214	0.215

Robust standard errors in parentheses, p values in italics
*** p<0.01, ** p<0.05, * p<0.1

Model : $\Delta NFP = \alpha_0 + \alpha_1 FS_{t-1} + \alpha_2 \Delta REF + \alpha_3 \Delta UFP + \alpha_4 \Delta SIZE + \sum_{l=2005}^{2006} \alpha_l YEAR_i + e_{it}$

$\Delta NFP_{it} = NFP_{it} - NFP_{it-1}$, the change in nonfinancial performance for Trust i in year t;

$\Delta REF_{it} = REF_{it} - REF_{it-1}$, the change in Trust cost efficiency measured by reference cost indices;

$\Delta UFP_{it} = UFP_{it} - UFP_{it-1}$,the change in underlying financial performance ;

$\Delta SIZE = SIZE_t - SIZE_{t-1}$, the change in the log of Trust total assets ;

$YEAR$ is an indicator variable coded 1 to 3 for the years 2004-2006 and e is an error term.

Table 24: Association between financial support and subsequent changes in nonfinancial performance (excluding specialist Trusts)

	Full sample	Reporting surplus	Reporting deficit
VARIABLES	∆NFP	∆NFP	∆NFP
	(1)	(2)	(3)
FS	0.149**	0.124*	0.284
	(0.080)	(0.068)	(0.253)
	0.067	*0.074*	*0.268*
∆REF	-0.105**	-0.056	-0.187*
	(0.048)	(0.045)	(0.099)
	0.033	*0.022*	*0.066*
∆UFP	-0.054	-0.048	-0.109
	(0.062)	(0.062)	(0.099)
	0.384	*0.445*	*0.278*
∆SIZE	-5.570	-1.291	-10.459
	(4.074)	(3.353)	(6.267)
	0.175	*0.702*	*0.103*
2005	2.629***	1.894**	-0.292
	(0.760)	(0.774)	(1.607)
	0.001	*0.018*	*0.857*
2006	omitted	omitted	-4.137
			(1.793)
			0.026
2007	3.542***	3.457***	Omitted
	(0.937)	(1.096)	
	0.000	*0.003*	
Constant	-2.843***	-2.667***	0.728
	(0666)	(0.654)	(1.550)
	0.000	*0.000*	*0.641*
Observations	152	88	64
R-squared	0.191	0.222	0.215

Robust standard errors in parentheses, p values in italics
*** p<0.01, ** p<0.05, * p<0.1

Model: $\Delta NFP = \alpha_0 + \alpha_1 FS_{t-1} + \alpha_2 \Delta REF + \alpha_3 \Delta UFP + \alpha_4 \Delta SIZE + \sum_{l=2005}^{2006} \alpha_l YEAR_i + e_{it}$

$\Delta NFP_{it} = NFP_{it} - NFP_{it-1}$, the change in nonfinancial performance for Trust i in year t ; $\Delta REF_{it} = REF_{it} - REF_{it-1}$, the change in Trust cost efficiency measured by reference cost indices; $\Delta UFP_{it} = UFP_{it} - UFP_{it-1}$, the change in underlying financial performance ; $\Delta SIZE = SIZE_t - SIZE_{t-1}$, the change in the log of Trust total assets ; $YEAR$ is an indicator variable coded 1 to 3 for the years 2004-2006 and e is an error term.

6.3.2.3 Research question 3: Was the receipt of financial support associated with an improvement in underlying financial performance in the subsequent period.

The provision of financial support was conditional upon a credible recovery plan which would bring the Trust back into financial balance without the need for further additional funding. As such the underlying financial performance of Trust in receipt of financial support should improve in the period subsequent to its receipt.

Tables 25 and 26 report the impact of financial support on subsequent financial performance for all Trusts receiving financial support (Table 25) and for general Trusts only (Table 26). In each Table three models are presented: Model (1) reports the results for all Trusts in the sample, Model (2) reports the results for Trusts which subsequently report a surplus, and Model (3) reports the results for Trusts which subsequently report a deficit.

Table 25 shows that, for the full sample (Model 1) the provision of financial support is strongly associated with a subsequent improvement in underlying financial performance (p= 0.002). The coefficient of 0.610 indicates that for a Trust receiving financial support of £5.7m underlying financial performance in the subsequent year will improve by £3.5m. This result is consistent with the requirement to produce a credible recovery plan to bring the Trust back into financial balance. Further, to the extent that that (unobserved) repayments of financial support took place, the improvement in underlying financial performance is understated. The results do however indicate that, on average, bringing a Trust back into financial balance takes more than one year. For Trusts reporting a surplus (Model 2) and Trusts remaining in deficit (Model 3) the results remain significant (p=0.007 and p=0.001 respectively) but the coefficient on Trusts reporting a surplus (0.355) is much lower than for Trusts remaining in deficit (1.160). The coefficient of 0.355 on Trusts reporting a surplus is associated with an improvement of £2.0m for an average sized Trust receiving an average £5.9m of financial support (see Table 19 Panel B). The mean underlying deficit for these Trusts is £5.4m (Table 19, Panel B) consistent with a period in excess of two years to achieve financial balance. The coefficient of 1.160 on Trusts remaining in deficit indicates that the subsequent underlying financial performance of a Trust receiving mean financial support of £5.3m would improve by £6.1m. However the much higher level of underlying deficit

for these Trusts (£11.9m, Table 19, Panel B) results in a subsequent underlying deficit which remains above £5m suggesting that these Trusts take about two years or more to move to financial breakeven. Table 26 shows the results for a sample which excludes specialist Trusts. As can be seen, the results are virtually the same. The coefficient on financial support is 0.627 (0.610) for the full sample, 0.360(0.355) for surplus Trusts and 1.160 (1.160) for deficit Trusts. As previously, all results were significant (p=0.004, p=0.017 and p= 0.001 respectively). Inclusion of the eight eliminated observations (untabulated) makes no difference to these results.

In summary, the provision of financial support is associated with significant improvements in financial performance in the subsequent year for all Trusts whether they reported deficits or surpluses. However, the improvement was bigger in Trusts which reported deficits than for those that reported surpluses. For these Trusts, the combination of a strong improvement in underlying financial performance alongside no improvement in nonfinancial performance suggests that underlying processes were not addressed and that, consistent with the Audit Commission view, financial support resulted in a strong focus on the scheduling of repayments, rather than acting as a spur to improve all aspects of performance.

Table 25: Association between financial support and subsequent changes in underlying financial performance

	Full sample	Reporting surplus	Reporting deficit
VARIABLES	ΔUFP	ΔUFP	ΔUFP
	(1)	(2)	(3)
FS	0.610***	0.355***	1.160***
	(0.195)	(0.127)	(0.326)
	0.002	*0.007*	*0.001*
ΔREF	-0.046	-0.003	-0.179
	(0.043)	(0.039)	(0.108)
	0.284	*0.949*	*0.103*
ΔNFP	-0.092	-0.053	-0.183
	(0.090)	(0.062)	(0.146)
	0.312	*0.402*	*0.216*
ΔSIZE	-6.110	3.234	-15.121
	(7.693)	(2.632)	(11.065)
	0.429	*0.224*	*0.179*
2005	-3.576***	-2.342***	-4.848**
	(0.925)	(0.682)	(1.805)
	0.000	*0.001*	*0.010*
2006	-5.188***	-2.247**	-8.638***
	(1.221)	(0.886)	(2.200)
	0.000	*0.014*	*0.000*
2007	omitted	omitted	omitted
Constant	-2.570***	1.099	3.735**
	(0.883)	(0.706)	(1.493)
	0.005	*0.125*	*0.016*
Observations	156	92	64
R-squared	0.314	0.250	0.506

Robust standard errors in parentheses, p values in italics
*** p<0.01, ** p<0.05, * p<0.1

Model: $\Delta UFP = \alpha_0 + \alpha_1 FS_{t-1} + \alpha_2 \Delta REF + \alpha_3 \Delta NFP + \alpha_4 \Delta SIZE + \sum_{l=2005}^{2006} \alpha_l YEAR_i + e_{it}$

$\Delta UFP_{it} = UFP_{it} - UFP_{it-1}$, the change in underlying financial performance ;

$\Delta NFP_{it} = NFP_{it} - NFP_{it-1}$, the change in nonfinancial performance for Trust i in year t ;

$\Delta REF_{it} = REF_{it} - REF_{it-1}$, the change in Trust cost efficiency measured by reference cost indices; $\Delta SIZE = SIZE_t - SIZE_{t-1}$, the change in the log of Trust total assets ; $YEAR$ is an indicator variable coded 1 to 3 for the years 2004-2006 and e is an error term.

Table 26: Association between financial support and subsequent changes in underlying financial performance (excluding specialist Trusts)

	Full sample	Reporting surplus	Reporting deficit
VARIABLES	ΔUFP	ΔUFP	ΔUFP
FS	0.627**	0.360**	1.160***
	(0.210)	(0.146)	(0.326)
	0.004	*0.017*	*0.001*
ΔREF	-0.041	-0.007	-0.179
	(0.044)	(0.043)	(0.108)
	0.356	*0.877*	*0.103*
ΔNFP	-0.092	-0.049	-0.183
	(0.090)	(0.063)	(0.146)
	0.309	*0.440*	*0.216*
ΔSIZE	-5.941	3.507	-15.121
	(7.701)	(2.644)	(11.065)
	0.443	*0.190*	*0.179*
2005	1.552	-2.224	-4.848**
	(1.162)	(0.632)	(1.805)
	0.185	*0.724*	*0.010*
2006	omitted	omitted	-8.638***
			(2.200)
			0.000
2007	5.238***	2.193**	Omitted
	(1.252)	(0.948)	
	0.000	*0.024*	
Constant	-2.644***	-1.127	3.735**
	(0.738)	(0.565)	(1.493)
	0.001	*0.051*	*0.016*
Observations	152	88	64
R-squared	0.313	0.235	0.506

Robust standard errors in parentheses, p values in italics
*** p<0.01, ** p<0.05, * p<0.1

Model : $\Delta UFP = \alpha_0 + \alpha_1 FS_{t-1} + \alpha_2 \Delta REF + \alpha_3 \Delta NFP + \alpha_4 \Delta SIZE + \sum_{l=2005}^{2006} \alpha_l YEAR_i + e_{it}$

$\Delta UFP_{it} = UFP_{it} - UFP_{it-1}$, the change in underlying financial performance ;

$\Delta NFP_{it} = NFP_{it} - NFP_{it-1}$, the change in nonfinancial performance for Trust i in year t ;

$\Delta REF_{it} = REF_{it} - REF_{it-1}$, the change in Trust cost efficiency measured by reference cost indices;

$\Delta SIZE = SIZE_t - SIZE_{t-1}$, the change in the log of Trust total assets ; $YEAR$ is an indicator variable coded 1 to 3 for the years 2004-2006 and e is an error term.

6.3.3 Study 2: Summary of findings

The system of financial support served to manipulate reported financial performance by shifting income between NHS organisations and to mislead key stakeholders such as Parliament and the public about the underlying financial performance of NHS Trusts. In the majority of cases the receipt of financial support by a Trust in financial difficulty served to advance the recognition of revenue from future periods as most financial support was provided on the basis that it would be 'repaid' through reductions in future years' revenue.

The policy of financial support had a significant impact on the accountability of NHS hospital Trusts. The amount of financial support in circulation in the period 2004-2006 amounted to £1bn and transformed a mean deficit of circa £8m to one of £2m for those Trusts receiving it. Further, of those Trusts which reported a surplus, 35% had an underlying deficit. There were also implications for the wider accountability of the NHS. The requirement to disclose financial support was restricted to Trusts receiving it in the year of receipt. There were no requirements for disclosure of subsequent 'repayments' of financial support nor of funds released for brokerage by Trusts in surplus. Further, although the system of financial support did not affect the financial performance of the NHS as a whole it did have an impact on wider NHS accountability by reducing the number of NHS Trust deficits and breaches of their statutory duty to breakeven reported annually to Parliament in the NHS Summarised Accounts. These findings are consistent with Thomas, Herrmann and Inoue, (2004) and Vinarri and Nasi (2008) who found that, in an institutional setting characterised by multi-layered accountability, manipulation of financial statements took place not only at the overall group level but also, through the manipulation of inter-group transactions, within the group, to meet individual entity accountability requirements.

The literature on the manipulation of financial performance however suggests that the flexibility accorded by mechanisms offering opportunities for manipulation can have beneficial consequences, provided that such flexibility is restricted to those who will work hard in the principals' interests and not afforded to those who will use the flexibility opportunistically. (Demski, Frimor and Sappington, 2004; Arya, Glover and Sunder, 2003; Scott, 2009). A second aspect of this study is

the investigation of the performance impact of the receipt of financial support for Trusts in financial difficulty.

The findings from this investigation indicate that, consistent with the policy's aims, financial support delivered benefits to patients and taxpayers in the form of better service standards whilst allowing time for financial difficulties to be addressed. The receipt of financial support is associated with a sizeable improvement in contemporaneous non-financial performance and a further, albeit smaller, improvement in the following year. Underlying financial performance also improves in the subsequent year. However, whilst, overall, the receipt of financial support was associated with improved service standards, the benefits were concentrated in those Trusts that reported a surplus. A significant minority of Trusts in receipt of financial support but reporting a deficit, (65/162=40%), demonstrated no improvement in service standards. So although overall, the system delivered benefits to service standards, these additional findings suggest that the system for selecting Trusts for receipt of financial support was not sufficiently discriminating or that the system was, in part, being used opportunistically to disguise deficits without the requisite effort being put into service standards.

Evidence in support of an opportunistic interpretation is provided by the results for the relationship between financial support and the following year's underlying financial performance. The receipt of financial support is associated with a significant improvement in the subsequent year's underlying financial performance not only for the full sample, but for both subsamples of Trusts reporting deficits as well as those reporting surpluses. The improvement in deficit Trusts however, was higher than in surplus Trusts, albeit against a background of higher mean deficits (£11.9m as compared with £5.4m). For these Trusts the significantly higher improvement in underlying financial performance combined with no improvement in nonfinancial performance is consistent with the Audit Commission's concern that, in failing Trusts, recovery planning focused 'to an unhealthy degree on the precise profiling of future borrowing and the schedule of repayments'. (Audit Commission, 2006b, para. 76) and is also consistent with an opportunistic interpretation that it was 'sometimes used to cover up deficits and then, at an appropriate time, the money was moved back again to where it had come from without any real effort being put into sorting out why the deficit arose in the other organisation first.' (House of Commons, 2006b: Q476).

In summary, the findings from this study indicate that the policy of income manipulation represented by financial support overall delivered benefits to patients and the public in the form of better service standards in Trusts in financial difficulty. However, in the context of a complex interaction of incentives and multi-layered accountability, there is also evidence that the system was not wholly effective and, in a significant proportion of Trusts, such benefits were not realised. In particular, in a demanding and evolving performance measurement regime, there were incentives for SHAs and PCTs to collude in the provision of financial support to Trusts in financial difficulty even where a robust recovery plan was not in place. Public and parliamentary scrutiny of the policy of financial support has, however, been hindered by the limited transparency associated with its impact on key measures of public accountability.

CHAPTER 7 CONCLUDING DISCUSSION

7.1 INTRODUCTION

NHS Trusts were introduced in 1991, as part of a wider programme of NPM reforms, with the aim of improving the performance, efficiency and accountability of NHS hospitals. Their establishment institutionalised a move away from a hierarchical mode of governance to a market based model. Operating as quasi-independent, self-governing organisations, they contracted with healthcare commissioners to deliver hospital services to patients. In this quasi-market it was envisaged that competition between service providers would drive up both the cost efficiency and quality of services. However, poorly developed costing systems undermined effective contracting and it was not until the early 21st century that sufficient investment was put into the development of the reference costing system which allowed the measurement of Trust cost efficiency and the pricing of treatments based on cost.

A key mechanism for delivering enhanced public accountability for NHS hospitals was the imposition of a statutory duty to breakeven, satisfaction of which was determined by reference to reported financial performance in the Trust's financial statements. During the 1990s this represented the main indicator by which Trust performance was assessed and, in the absence of other, more robust, indicators, also acted as a proxy for cost efficiency: a Trust achieving financial balance was assumed to be operating efficiently. However, concerns that the focus on financial performance was having an adverse impact on service standards led, in 2002, to the introduction of a multi-dimensional performance measurement system, the Star ratings, aimed at incentivising better service standards. In turn, the Star ratings system was subject to criticism as a result of its perceived vulnerability to manipulation and in 2005-06 it was replaced by the Annual Health Check which was characterised by wider measurement diversity and a more even balance in the weightings attached to financial and nonfinancial measures.

The statutory duty to breakeven, however, remained the primary measure for public accountability purposes and considerable import continued to be accorded to reported financial performance throughout the period of study. In the context of

rising demand, incomplete contracting, and the natural volatility of some revenue expenses this breakeven requirement became increasingly challenging. However, consistent with the slow response of hospital accounting systems to the move away from hierarchical modes of governance, a number of mechanisms which offered flexibility in the reporting of Trust financial performance were initially retained. These flexibilities, such as the ability to transfer resources from the capital budget to the revenue account and the ability to defer patient treatment into a future period, were gradually withdrawn in response to new NPM initiatives. During the period of this study therefore only one major source of flexibility remained. This was the system of financial support, in which surplus funds from elsewhere in the NHS were transferred, as non-recurrent additional revenue, into Trusts in financial difficulty. The provision of financial support was conditional upon an agreed recovery plan and was often provided in the expectation that anticipated future savings would accommodate 'repayment' of the funds advanced. Thus financial support acted as a form of income shifting within the NHS as a whole and as a form of accelerated revenue recognition in Trusts receiving it.

The conventional view of financial support was that it was provided to Trusts which were in financial difficulty to avoid the adverse consequences for patient care of short term cost cutting to achieve financial breakeven. Over the period of this study, however, concern was growing that financial support was being applied opportunistically to disguise deficits rather than to address underlying issues. These concerns were exacerbated by the lack of transparency concerning the impact of financial support on reported financial performance. All financial support transactions, for both receipt and repayment, were effected through adjustments to the revenue account and, until 2002-03, were completely hidden from public scrutiny. After 2002-03 concerns raised by the Audit Commission led to the disclosure of financial support in the financial statements of Trusts in financial difficulty. However, disclosure was limited to Trusts in receipt of financial support in the year of receipt. There were no similar disclosure requirements for 'repayment' or for those Trusts who released their surpluses for brokerage.

7.2 RESEARCH INVESTIGATIONS

The research paradigm adopted in this thesis is positivist and an empirical archival methodology is applied, using an agency perspective to motivate the

generation of research questions. This approach, whilst recognising the specific institutional and incentive framework within the NHS, has been facilitated by the increased emphasis on quasi-market mechanisms to address perceived issues of performance and efficiency in NHS hospitals. Two main studies have been undertaken. Study 1 investigates the relationship between service standards and cost efficiency in English NHS hospitals over the period 2002-03 to 2007-08 and the relative impact on this relationship of the AHC as compared with the Star ratings. Study 2 investigates the performance and accountability impact of the flexibility in reported financial performance afforded by the system of financial support to Trusts in financial difficulty over the period of financial support disclosures from 2003-04 to 2005-06.

Data to facilitate these investigations is drawn from public sources. Data on Trust cost efficiency, in the form of Trust reference cost indices, is obtained from the Department of Health; data on Trust financial performance is obtained from the Laing and Buisson database of NHS financial reports and data on nonfinancial performance, from which a uniform measure of nonfinancial performance *(NFP)* is constructed, is obtained from the Healthcare Commission and its predecessor, the Commission for Health Improvement.

A key element of the investigations in this thesis is the formulation of a stylised objective function for NHS Trusts which draws, for its authority, on key Government policy documents to model performance against service standards as a function of cost efficiency in the presence of a financial resource constraint determined by Trusts' statutory duty to breakeven. This objective function then serves as a basis for the application of production possibility frontier analysis aimed at the generation of predictions for the impact of PMS design on the incentivisation of cost efficiency as a means of improving service standards. Particular reference is made to the potential impact of, first, the balance between financial and nonfinancial measures and, second, measurement diversity in the form of coverage of activities, and the scope that these features offer for manipulation of the system.

A further consideration in this stage of the analysis is the impact of PMS design on the incentive to manipulate reported financial performance as a means of improving performance ratings. Two forms of this type of manipulation which have a potential impact on Trust efficiency are considered in this thesis: first the trading off

of financial for nonfinancial performance and secondly, the selective use of an NHS specific transaction, financial support, which served to shift income from NHS organisations in surplus to those in financial difficulty and, in Trusts receiving support, to accelerate the recognition of revenue.

Study 2 also draws on the stylised objective function and the same measures of performance to investigate whether the flexibility afforded to reported financial performance by the accounting treatment of financial support generated benefits in the form of better performance in NHS Trusts in financial difficulty. This question is broken down into three subsidiary questions. First, did the receipt of financial support lead to an improvement in nonfinancial performance (*NFP*) in the year of receipt? Second, did benefits in the form of further improvements in nonfinancial performance continue to flow through in the year subsequent to receipt? Lastly, consistent with the aims of the policy of financial support, which was to allow Trusts time to address financial difficulties, was the receipt of financial support associated with an improvement in underlying financial performance in the year following receipt?

7.3 FINDINGS

The findings from Study 1 indicate that the relationship between performance against Government measures and cost efficiency is significantly positive whichever performance measurement system is in place. This finding is consistent with the findings of previous studies in the frontier analysis literature (Ludwig, Van Merode and Groot, 2010; Deily and McKay, 2006) and with the findings of studies such as Propper, Sutton, Whitnall and Windjmeier (2010) where better quality of care in English hospitals as compared with Scottish hospitals was consistent with process improvements implemented in order to meet Government targets. Nonparametric analysis however showed that the trading off of financial for nonfinancial performance was greater in the Star ratings than in the AHC, consistent with the Star ratings being more vulnerable to manipulation. Further, results from multivariate analysis indicate that the relationship between nonfinancial performance and cost efficiency in the AHC is significantly higher than in the Star ratings, consistent with reduced potential for manipulation. This result is robust to choice of financial performance measure and to the exclusion of the best performing Trusts, those which converted to Foundation Trust status during the period. This

finding is of particular interest because it provides evidence of the impact of PMS design on less well performing and inefficient Trusts. Although the inclusion of specialist Trusts, with very different performance characteristics, results in a loss of significance in the difference between the two performance systems, the results clearly point to the AHC as being a more effective vehicle for incentivising cost efficiency as a means of improving nonfinancial performance, than the Star ratings.

Notwithstanding the findings from Study 1 which suggest that reduced potential for PMS manipulation facilitates the incentivisation of cost efficiency whilst maintaining service standards, the findings from Study 2 suggest that, for Trusts in financial difficulty, some flexibility in the measurement of financial performance can be beneficial. The findings from Study 2 provide evidence that, over the whole sample, the policy of financial support generated benefits in the form of improved service standards not only in the year of receipt but also in the following year. The level of nonfinancial performance for Trusts in receipt of financial support was significantly higher than the performance adjusted benchmark level of nonfinancial performance, based on the performance of a reference group of Trusts not in receipt of financial support and further on-going improvements were observed in the following year. Further, consistent with the objective of allowing Trusts time to address their financial difficulties, subsequent improvements in underlying financial performance were also observed. However, when the full sample was segregated, into those Trusts subsequently reporting a surplus and those reporting a deficit, significant differences arose. Although improvements in underlying financial performance were observed in both sub-samples, the benefits to service standards were found to be concentrated in those Trusts which subsequently reported a surplus. The significant minority of Trusts (40%) reporting a deficit showed no improvement in service standards. These findings are consistent with the Audit Commission's concerns that in failing Trusts undue emphasis was placed on the repayment of financial support rather than on addressing underlying issues (Audit Commission, 2006b). So although, overall, the system delivered benefits in the form of improved service standards, these additional findings suggest that the system for selecting Trusts for the receipt of financial support was not sufficiently robust or that, in an increasingly demanding performance regime and multi-layered accountability, an evolving incentive framework led to the opportunistic application of financial

support to disguise deficits without the requisite effort being put into performance improvements.

The accounting policy for financial support had a significant impact on the accountability of NHS Trusts. The amount of financial support in circulation in the period 2004-2006 amounted to £1bn and transformed a mean deficit of circa £8m to one of £2m for those Trusts receiving it. Further, of those Trusts which reported a surplus, 35% had an underlying deficit. So, although the system of financial support did not affect the reported financial performance of the NHS as a whole, it reduced the number of NHS Trusts in deficit, and in breach of the statutory duty reported annually to Parliament by the Department of Health in the NHS Summarised Accounts. There were also other implications for accountability. The requirement to disclose financial support, which was introduced in 2002-03, was restricted to Trusts receiving support, and only in the year of receipt. There were no similar disclosure requirements for subsequent 'repayments' of financial support or of funds released for brokerage by Trusts in surplus.

7.4 CONTRIBUTION

This thesis makes the following contributions to the literature. First, it makes a contribution to the limited number of empirical archival studies in public sector accounting research in the UK, Europe and Australasia, particularly in the domain of performance measurement where a qualitative paradigm is dominant. Secondly, it makes a contribution to the literature on the cost efficiency of hospitals by utilising broad based measures of cost efficiency and service standards to investigate the relationship between them in the presence of multi-dimensional performance measurement systems. Thirdly, it makes a contribution to the literature on public sector performance measurement, including the balanced scorecard literature, by investigating, in a quasi-experimental setting, the impact of PMS design on the incentivisation of cost efficiency as a means of improving service standards. Fourthly, it makes a contribution to the public sector literature on the manipulation of financial performance, by investigating whether a form of manipulation, as manifest by the accounting policy for financial support in English acute hospitals, can generate benefits in the form of better performance. Fifthly, it makes a contribution to the wider literature on the manipulation of financial performance, by investigating manipulation in a novel setting characterised by a multiple layers of accountability

and interacting incentives. Finally, it makes a contribution to the literature on financial accountability and the application of UK GAAP in the UK public sector by investigating the impact of financial support on the financial accountability not only of NHS Trusts, but also of the wider NHS.

7.5 POLICY DISCUSSION

There are a number of policy implications arising from the findings of this thesis.

First, in the context of a multi-dimensional performance measurement system, the findings from Study 1 add to previous findings that cost efficiency and service standards are positively associated. However, although a key objective of the two performance measurement systems was to incentivise cost efficiency as a means of improving service standards, neither system incorporated specific measures of cost efficiency as was originally intended (Department of Health, 1999, 2000) and as is characteristic of other hospital systems, such as the Ontario Hospital Report in Canada. Measures which have typically been adopted in Canada for example include not only measures related to costs and liquidity but also to aspects of process efficiency such as theatre utilisation rates. Although reference costs and Trust reference cost indices were initially subject to criticism because of data quality and coding issues, the reference costing system has now been in operation for over ten years, with substantial investment to address early concerns. Its outputs are now used by a wide range of users for a variety of purposes, including allocation of resources to commissioners and contracting between commissioners and providers. Reported financial performance, which has typically been used as a proxy for cost efficiency, is an imprecise measure which could beneficially be supplemented by other measures of efficiency, including the Trust reference cost index.

The positive association of service standards and cost efficiency also provides evidence that the multi-dimensional performance measurement systems introduced by the Labour Government do provide a framework within which service standards are not, overall, sacrificed in the pursuit of cost efficiency. Study 1 further indicated that the AHC, characterised by wider measurement diversity and a better balance between financial and nonfinancial measures, incentivised cost efficiency

as a means of improving service standards, more effectively than the Star ratings. These findings are consistent with the interpretation of the results of other studies not directly investigating cost efficiency and service standards which suggest that achievement of waiting targets is associated with improvements in other, unmeasured aspects of care, consistent with the improvement of business processes that generate benefits beyond immediate target achievement. The findings from the study in this thesis, that the AHC more strongly incentivises improvements in service standards through improved cost efficiency suggests that, in a review of PMS design, such as that currently being undertaken by the Care Quality Commission[51], care should be exercised in reducing measurement diversity and in weakening the incentives to achieve good performance through improved efficiency. Initial reports[52] suggest that, in the future, waiting targets may not feature as strongly in a new performance measurement system. Performance standards are to become advisory rather than mandated and, with predictions that Trusts offering lower quality services will attract lower funding, market forces are to be relied upon to maintain service standards and promote efficiency. Such a system would be similar to that envisaged in the 1980s. However, policy documents have not yet made clear what the implications of such a system would be for NHS Trusts and local service provision, as compared with the current targets based system[53]. The empirical archival research findings for the current system of mandated targets and standards suggest that, overall, such a system incentivised local hospitals to improve their services cost efficiently within their resource constraints, through process improvements which have produced wider benefits for overall quality of services. A system which relies on market forces may, in contrast, result in the gradual withdrawal of services from some local communities.

Whilst acknowledging that the potential for gaming and manipulation which characterises any performance measurement system, can have adverse consequences for patients, the evidence from the research literature which focuses

[51] The Care Quality Commission replaced the Healthcare Commission from 1 April 2009.
[52] http://www.bbc.co.uk/news/10469270
[53] A number of broad policy documents have been issued by the Care Quality Commission but performance assessments of hospital Trusts will not be provided for 2009/10 and, at the present time, there are no published policy documents for the assessment of acute services. http://www.cqc.org.uk/guidanceforprofessionals/nhstrusts.cfm

on the adverse consequences of the Star ratings and the AHC is largely restricted to critical and case study approaches where the findings are more difficult to generalise (see for example, Bevan and Hood, 2006). The non-parametric analysis of Study 1 however provides evidence that financial breakeven was the target most frequently sacrificed, albeit less so in the AHC. However, whilst some scepticism may be expressed regarding the ability of market based mechanisms to deliver improvements in care, in the absence of further information about proposed developments and in the light of the findings from Study 1, it may be that the Care Quality Commission's review will represent a further step in Courty and Marschke's (2007) evolution of performance measurement in response to revealed gaming which will result in improved goal congruence and reduced manipulation.

Whereas NPM policy initiatives in the early 21st century were consistent with a negative view of the manipulation of service standards to meet accountability objectives, a more tolerant view was taken of the manipulation of financial performance where service standards would otherwise be jeopardised. During the period of this thesis one particular source of flexibility in reported financial performance, the system of financial support, was available to NHS Trusts in financial difficulty. The findings from this study indicate that, for these Trusts, this policy of income manipulation delivered overall benefits to patients and the public in the form of better service standards. The finding of an improvement in service standards in Trusts receiving financial support as compared with those that did not suggests that there is an argument for the retention of such a policy in a demanding performance measurement environment where the achievement of accountability objectives, such as the statutory duty to breakeven, may involve managerial action which is potentially damaging to standards of care.

However, Study 2 also provided evidence consistent with some opportunistic application of financial support, which may have been facilitated by the limited transparency around financial support transactions. The accounting policy for financial support, significantly impaired Trust and wider NHS accountability and, in the context of multi-layered accountability allowed Trusts, SHAs and PCTs to cooperate to disguise deficits without addressing underlying issues. These findings raise interesting questions related to the withdrawal, in 2006-07, of the flexibility in reported financial performance which financial support provided. The withdrawal of financial support addressed the issue of accountability but at the cost of losing the

potential benefits of flexibility in reported financial performance. Rather than abandoning the system altogether, an alternative solution might have been to improve the transparency of financial support transactions, by requiring more comprehensive disclosures, thereby allowing better public scrutiny of the policy.

The continuing programme of NPM reforms in the NHS has, over the last thirty years, resulted in tensions between the 'new' market modes of governance and the previous hierarchical mode of governance. In this context, the investigation of financial support also exposes some interesting issues concerning the nature of the transaction, the accounting policy that was adopted and the interaction of Trust accountability and wider NHS accountability. The nature of financial support transactions was ambiguous, consistent with the slow adaptation of NHS accounting systems to the new marketised NPM environment. From a conventional GAAP perspective, financial support transactions had, variously, the characteristics of loans (where support was repayable), non-recurring 'grants' or 'non-reciprocal transfers' (where support was not repayable), income shifting (internal manipulation to generate added value), and accelerated revenue recognition (earnings management). Further, from a hierarchical perspective, financial support might be viewed as a budget reallocation, consistent with hospitals' previous status as cost centres within a local health authority. A more refined consideration of the nature of such transactions would have facilitated the development of disclosure requirements which recognised and acknowledged the public sector setting whilst realising the benefits which might accrue from the selective manipulation of reported financial performance. Consideration, for example, could have been given to more clearly identifying the nature of the transactions, and more rigorous disclosure requirements capturing not only the receipt of financial support but also its repayment, with equivalent disclosures being made in the financial statements of NHS organisations releasing surpluses for brokerage.

In 2006-07, however, the flexibility which was accorded to reported financial performance by the system of financial support was withdrawn. The arguments in favour of abandoning financial support included greater transparency in financial reporting and an equitable system of funding for a diverse range of providers of acute hospital services. However, NHS Trusts operate under more onerous constraints than other types of provider. They not only have a statutory duty to breakeven but their ability to adjust their service portfolio is constrained by the

requirements associated with public consultation, which can prevent service adjustments being made even in the presence of supporting clinical arguments. In the context of more onerous constraints on their service portfolio, there is an argument that NHS Trusts should be treated differently to other types of provider and that some flexibility in reported financial performance is appropriate.

A system of interest bearing loans is now in place for NHS Trusts in financial difficulty which permits liquidity issues to be addressed and contingency reserves are held centrally in order to secure overall NHS financial balance. This system does not, however, allow any flexibility in reported financial performance and thus does not address the incentive to engage in more damaging responses to the breakeven target. Indeed, the carrying of surpluses and of contingency reserves to deliver overall NHS balance may, from a public accountability perspective, be criticised as not generating the best outcomes for patients within a resource constrained environment. There is an argument therefore that, in the context of evidence that the recirculation of surpluses within the NHS delivers benefits to patients, the maintenance of such reserves should be subject to more explicit scrutiny and be justified on service delivery grounds.

During the period of this study, the best performing Trusts have been permitted to transfer to Foundation Trust status. Foundation Trusts, which are excluded from this study, are subject to a different model of regulation. The Labour Government's vision was that, with its programme of reforms in place, all NHS Trusts would be in a position to become Foundation Trusts by 2008. During the period of this study however the majority of Trusts remained as NHS Trusts and a significant number of NHS Trusts are yet to convert. This raises questions about the robustness of the remaining Trusts and whether Foundation Trust status is appropriate for all hospitals. For those which remain as NHS Trusts there is an argument, on the basis of the findings here, for considering the reinstatement of the flexibility that the policy of financial support afforded hospitals in managing the balance between service standards and financial performance.

7.6 LIMITATIONS

A number of limitations characterise the studies in this thesis. Those related to measurement and data issues are considered in this section whilst those

limitations which are associated with opportunities for further research are considered in the next section, (Section 7.7).

7.6.1 Nonfinancial performance

The stylised objective function which models service quality, in the presence of a financial resource constraint, as a function of cost efficiency underpins much of the analysis in this thesis. In order to operationalise this function, a uniform measure of nonfinancial performance (*NFP*) is constructed from data on Trust performance against key Government measures. These measures were identified through public and professional consultation as being the primary concerns of the public and so represent a proxy for alignment with stakeholder interests. They are dominated by access measures, in the form of waiting times for treatment. As waiting times often influence clinical outcomes, with many conditions deteriorating the longer the wait for treatment, and adverse consequences for the ultimate outcome, waiting times do represent a proxy for the quality of treatment, as well as representing a concern of patients whose quality of life may be impaired during the period of wait.

However they do not measure the quality of care once treatment is undertaken. This aspect of service standards is regulated by the medical profession, and includes key indicators such as mortality, readmission and complication rates. *NFP* does not include such clinical measures of quality although some measures do provide a context for the analysis of the findings from this study. Mortality and readmission rates are, for example, incorporated into the 'balanced scorecard' component of the Star ratings system and in the AHC, which is dominated by process indicators, is implicitly captured by indicators such as the 'standard of patient protection systems', 'clinical governance', 'audit participation' and 'MRSA' indicators. These additional measures provide part of the context for the investigations in this thesis but, because of lack of comparability, they are not incorporated into the aggregate measure of nonfinancial performance *NFP*.

Detailed data is available on indicators such as mortality and readmission rates are available and these have been used in other studies, largely in the economics literature, which investigate the impact of performance measurement systems on quality of care. However, these measures represent a very narrow measure of quality of care and, in the absence of a consultation exercise of the kind

that characterised the development of both the Star ratings and the Annual Health Check, cannot easily be aggregated with other measures to form a composite measure of quality. Further, although appearing to be a more direct measure of service standards, use of these measures is not without problems as they need to be adjusted for case mix if inappropriate policy conclusions are to be avoided. The dangers associated with a crude interpretation of mortality rates, for example, include a retreat to defensive medicine with its consequent impact on the skills of medical staff, the development of new and innovative treatments and perversely, on the clinical outcomes for high risk patients.

Although *NFP* does not directly incorporate measures of clinical quality, such as mortality rates, evidence that it represents a reasonable proxy for these aspects of care is provided by studies in the economics literature which demonstrate that performance against waiting targets is positively associated with other quality of care indicators such as mortality, morbidity and re-admission rates. In this respect waiting targets, although narrow in themselves, can be considered, consistent with Bevan and Hood's (2006) concept of synecdoche, as being representative of the whole.

7.6.2 Trust cost efficiency

The measure for cost efficiency used in this thesis is the Trust reference cost index. This is a relative measure of efficiency and so, although the multivariate analysis can identify the relationship between relative efficiency and the level of nonfinancial performance (*NFP*) the findings say nothing about the absolute levels of efficiency and the trends in absolute efficiency over time. This however was not the remit of this thesis and other studies, such as Phelps, (2009) and LeGrand (2002a) have sought to do this using alternative methods. Other measures of efficiency such as theatre utilisation are used in other international PMS. However, these suffer from similar limitations to the use of narrow measures of quality such as mortality rates: they are a narrow measure of efficiency, covering only surgery, not medicine or diagnostic services, and increased theatre utilisation may have adverse consequences for patient safety. Further, the data for these alternative measures are not readily available and are not subject to data control and quality checks in the same way as reference costs. Further, there is no means of aggregating them into a

broad based, consistent measure of efficiency such as is represented by the Trust reference cost index.

Alternative measures of Trust cost efficiency can be derived using frontier analysis techniques but these techniques have been subject to sustained criticism for the sensitivity of results to model specification and choice of input and output variables. In the NHS the use of these techniques could be further criticised for using data which is subject to less quality control than the underlying data used for constructing reference cost indices.

7.6.3 Financial support

The aim of Study 2 is to investigate the performance impact of the receipt of financial support for Trusts in financial difficulty. Data for financial support was available only from 2003-04 to 2005-06, (after which the system was withdrawn). The period of study was therefore restricted to the three years of disclosure and by necessity excluded 2002-03 for which data on efficiency and nonfinancial performance was available and which featured in the investigations for Study 1. Further, disclosures for financial support were only required for financial support received. Repayments and the release of funds for brokerage were not subject to disclosure, nor was it possible to identify the Trusts where the financial support was provided on condition that it was repayable. Information from Audit Commission reports and the descriptive statistics in Section 7.6.3 that over £900m of financial support was provided to NHS Trusts over the period of study, and that therefore the majority, at least two thirds, of financial support was provided a repayable basis.

An underlying assumption of Study 2 is that the funds transferred into NHS Trusts in financial difficulty were, temporarily, genuinely surplus to requirements and that service standards elsewhere in the NHS were not adversely affected by a temporary relinquishment of them. It has not been possible to test this assumption because of the lack of availability of data on the release of funds for brokerage. However, transfers of funds out of NHS organisations with surpluses, whether SHAs, PCTs or NHS Trusts, are subject to agreement by their Boards and the assumption that these organisations would be unlikely to agree such a transfer if it were to adversely affect their own service standards is not an unreasonable one

given the demanding performance environment that the Star ratings and the Annual Health Check represented.

7.7 FURTHER RESEARCH

The findings from this thesis suggest a number of opportunities for further research.

Study 1 found that the AHC was more effective at incentivising cost efficiency as a means of improving service standards but it has not been possible to identify which features of the AHC had the most impact on this increased effectiveness. However, further investigation of this question may be possible using alternative methods such as principal component analysis. The findings of such a study would further contribute to our understanding of PMS design in a healthcare setting and could serve to inform future PMS design in the NHS. Further, although the AHC was found to be more effective than the Star ratings, there has been no assessment in this thesis of the cost of operating these systems. An assessment of these costs would permit a cost-benefit analysis which could further inform future PMS design.

Study 2 investigates the performance impact of financial support for Trusts in financial difficulty and finds that, overall, the system served to improve performance. In 2006-07 the system of financial support and the flexibility it accorded in the reporting of financial performance was withdrawn. A further study of Trusts in financial difficulty in this new regime could offer further insight into the incentive to engage in short term cost cutting to meet the financial breakeven target which may have adverse consequences for service standards. An empirical archival study drawing on the literature on the manipulation of real transactions to achieve financial objectives may prove fruitful although the challenges of developing an appropriate research design in the context of the complex incentive framework of the NHS would be substantial. A qualitative research paradigm and an interview based method, may however offer an alternative approach to this question. The findings from such a study would add to our understanding of the manipulation of financial performance in a public sector setting and would have the potential for informing future accounting policy in the public sector.

This thesis is restricted to a consideration of NHS acute Trusts. However the full sample includes a number of specialist Trusts whose performance characteristics are shown by univariate analysis to be significantly different from acute general hospitals. Inclusion of specialist Trusts in the sample reduces the significance of the findings in some cases, particularly for Study 1, which investigates the relative impact of the AHC as compared with the Stars on the incentivisation of cost efficiency. However the number of specialist Trust observations is insufficient to conduct the investigations in this thesis on a sample of specialist Trusts only. The performance of these Trusts is however of interest. They are observed to have better financial performance, similar nonfinancial performance but much worse cost efficiency than acute general Trusts. Further insight into these smaller Trusts and the impact that the imposition of a performance measurement system on them as compared with general hospitals may however be possible within the context of an alternative research paradigm and method. Such a study has the potential for shedding light on the reasons for different performance characteristics of specialist Trusts which could have relevance for the development of future service design and delivery.

Foundation Trusts, which began to be established in the latter part of the period of study, and which are subject to a different form of financial regulation, have been excluded from the studies in this thesis. Foundation Trusts offer a different setting in which to explore the impact of performance measurement. As they are not subject to the statutory duty to break-even they also offer the possibility of a quasi-experimental study which explores the impact of the financial breakeven duty on the manipulation of financial performance by comparing a sample of Foundation Trusts with a sample of NHS Trusts.

Finally the research questions in this thesis have been approached using a positivist research paradigm and an empirical archival approach facilitated by predictions developed from an NPM informed agency perspective. This approach has been facilitated by the increasingly marketised environment in which NHS hospital services are delivered. The public sector literature which adopts a qualitative paradigm is extensive but the dominance of this paradigm potentially restricts the development of the wider public sector research agenda. Consistent with Goddard (2010) the findings from this thesis have suggested a number of ways in which alternative paradigms and models can further develop and add to our

understanding of performance measurement systems and of mechanisms which accord flexibility in the reporting of financial performance in public sector organisations.

7.8 PERSONAL REFLECTION

There were numerous challenges associated with the development of the studies in this thesis. My primary interest was the incentive impact of balanced scorecard inspired performance measurement and management and I had data on both financial and nonfinancial performance from the Laing and Buisson database and from the Healthcare Commission.

The first challenge was the development of a broad measure of service standards. I had very broad aggregate measures of performance in terms of Trust performance ratings but the performance ratings under the Star system were not directly comparable with those under the Annual Health Check. The first breakthrough was that, although the Star ratings and the AHC each had a number of distinctive features, there was a degree of commonality in the key performance targets which were determined by the Government, after consultation with the public, as the main priorities for the NHS. It was these similarities which permitted the construction of a uniform aggregate measure of service standards (*NFP*) that opened the way to investigation of the impact of performance measurement systems.

The balanced scorecard empirical archival literature is almost wholly focused on the private sector. My initial research designs were based on this literature and focused on reported financial performance. Initially, none of the analysis I performed gave consistent results. A second breakthrough came after an extensive reading of policy documents. This led to two developments: first, the acknowledgement that, in the public sector, reported financial performance represented a poor measure of 'financial performance' as it was effectively a measure of the ability of the organisation to deliver services within a resource constraint. Second, as a consequence, cost efficiency, for which data were available, was a better measure of 'financial performance' in the NHS context and was consistent with Government policy objectives. These two insights led to the final major breakthrough which was to formulate an objective function for NHS

Trusts based on Government policy statements which models service standards as a function of cost efficiency and the financial breakeven requirement.

The development of *NFP*, (an aggregate measure of service standards), the availability of Trust reference cost indices as measures of Trust cost efficiency and the formulation of a Trust objective function have not only facilitated the investigation of the research questions in this thesis but have also opened up opportunities for the investigation of other research questions within the healthcare sector and, in terms of modelling, in the wider public sector.

A further hurdle to overcome in the analysis of performance measurement was the identification of 'financial support' as a form of manipulation which served to disguise the underlying financial performance of Trusts. Given the largely undisclosed nature of these transactions and the ambiguity of language used in official documents, the identification of the nature of financial support transactions proved particularly challenging and required persistence, ingenuity and long periods of reflection to synthesise and digest information from a diverse range of sources. This work however eventually provided a basis to explicitly link the policy of financial support, which permitted flexibility in accounting for revenue, to propositions in the formal agency-based literature that manipulation of financial performance can be beneficial if measures taken to address gaming result in other, more damaging, consequences for service delivery. The investigations in this thesis have formed the basis of four papers submitted for publication in internationally recognised journals.

The experience of developing this thesis highlights the importance, particularly in the context of public sector research, of diligent and careful study of policy documents. Persistence and determination in the identification of features distinctive to the institutional setting facilitates a deeper understanding of the impact of the regulatory framework and of organisational design on managerial incentives. This understanding, in turn, is a necessary pre-requisite for the identification and motivation of innovative, well-designed research questions and methods, for the interpretation of the findings and for the generation of appropriate policy recommendations.

CHAPTER 8 BIBILOGRAPHY

Agrizzi, D. (2008). 'Assessing English hospitals: contradiction and conflict'. *Journal of Accounting & Organizational Change*, 4(3): 222-242.

Ahmed A., Takeda S.C., Thomas S. (1999). 'Bank loan loss provisions: a re-examination of capital management, earnings management and signaling effects'. *Journal of Accounting and Economics*, 28:1–25.

Aidemark, L-G. (2001). 'The meaning of balanced scorecards in the healthcare organisation'. *Financial Accountability and Management*, 17(1): 23-39.

Alali, F. and Jaggi, B. (2011). 'Earnings versus capital ratios management: role of bank types and SFAS 114'. *Review of quantitative finance and accounting,* 36:105-132.

Amat, C. and Gowthorpe, O. (2005). 'Creative accounting: Some ethical issues of macro- and micro-manipulation'. *Journal of Business Ethics*, 57:55-64.

Arah, O. A., Klazinga, N. S., Delnoij, D. M. J., Ten Asbroek, A. H. A., Custers, T. (2003). 'Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement'. *International Journal for Quality in Health Care, 15*(5): 377–398.

Argyris, C. (1977), 'Organizational learning and management information systems'. *Accounting, Organizations and Society*, 2(2):113-123.

Arya, A., Glover, J. C. and Sunder, S. (2003).'Are Unmanaged Earnings Always Better for Shareholders?' *Accounting Horizons,* Supplement: 111-116.

Audit Commission (2004). *Introducing Payment by Results: Getting the Balance Right for the NHS and Taxpayers*. HAR3237.

Audit Commission (2005). *Financial Management in the NHS: NHS (England) Summarised Accounts 2003-04*. The Stationery Office Limited, HC60-1 179531 06/05 77240.

Audit Commission (2006a). *Financial Management in the NHS: NHS (England) Summarised Accounts 2004-05*. HC1092-I, The Stationery Office.

Audit Commission (2006b). *Learning the lessons of financial failure in the NHS*. HNR3352 ISBN 1-86240-524-7.

Audit Commission (2010a). '*Reference Costs – Review of Uses by NHS Bodies'*, Gateway reference: 14618.

Audit Commission (2010b). '*Review of Reference Costs- Findings from pilot reviews of the quality of 2008/09 submissions*', Gateway reference: 14618.

Bailey, E. E. (1981). '*Contestability and the Design of Regulation and Anti-trust Policy*', *American Economic Review,* Papers and Proceedings:178-183.

Ballantine, J., Brignall, S. and Modell, S. (1998). 'Performance measurement and management in public health services: a comparison of U.K. and Swedish practice'. *Management Accounting Research*, 9:71-94.

Ballantine, J., Forker J. and Greenwood, M. (2007). 'Earnings management in English NHS hospital trusts'. *Financial Accountability and Management*, 23(4): 421-440.

Ballantine, J., Forker J. and Greenwood, M. (2008a). 'Public and private sector auditors and accruals quality in English NHS hospital Trusts'. *British Accounting Review,* 40:28-47.

Ballantine, J., Forker J. and Greenwood, M. (2008b). 'The governance of CEO incentives in English NHS hospital trusts'. *Financial Accountability and Management*, 24(4): 385-410.

Banker, R.D and Datar, S.M. (1989). 'Sensitivity, precision and linear aggregation of signals for performance evaluation' *Journal of Accounting Research* 27(1):21-39.

Banker, R.D., Potter, G. and Srinivasan, D. (2000). 'An empirical investigation of an incentive plan that includes non-financial performance measures. *The Accounting Review*, 75(1):65–93.

Bar-Gill, O. and Bebchuk, L. (2003). 'Misreporting corporate performance'. Working paper. Harvard University.

Barton, A. D. (1999). 'Public and Private Sector Accounting—the Non-Identical Twins', *Australian Accounting Review,* 9(2):22–31.

Barton, A. D. (2000). 'Accounting for Public Heritage Facilities—Assets or Liabilities of the Government?' *Accounting, Auditing and Accountability Journal*, 13(2): 219-236.

Barton, A. D. (2004). 'How to Profit From Defence: A Study of the Misapplication of Business Accounting to the Public Sector in Australia', *Financial Accountability and Management*, 20(3):281-304.

Barton, A. D. (2005). 'Professional Accounting Standards and the Public Sector—a Mismatch'. *Abacus*, 41(2):138-158.

Basioudis, I.J. and Ellwood, S. (2005a). 'An empirical investigation of price competition and industry specialisation in NHS audit services'. *Financial Accountability and Management,* 21(2): 219-248.

Basioudis, I.J. and Ellwood, S. (2005b). 'External audit in the National Health Service in England and Wales: A study of an oversight body's control of auditor remuneration'. *Journal of Accounting and Public Policy,* 24:207–241.

Baumol, W.J. (1982). *'*Contestable Markets: An Uprising in the Theory of Industry Structure'. *American Economic Review*, 72:1-15.

Baumol, J.C., Panzar, J.C. and Willig, R.D. (1982). '*Contestable Markets and the Theory of Industry Structure'*. San Diego: Harcourt Brace Jovanovich.

Beatty, A., Chamberlain S.L. and Magliolo, J. (1995). 'Managing financial reports of commercial banks: the influence of taxes, regulatory capital, and earnings'. *Journal of Accounting Research* 33:231–261.

Beaver, W., McNichols, M. and. Nelson K. (2007). 'An Alternative Interpretation of the Discontinuity in Earnings Distributions'. *Review of Accounting Studies* 12: 525–56.

Beneish, M. (1997). 'Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance'. *Journal of Accounting and Public Policy,* 16: 271-309.

Bessire, D. and Baker, C.R. (2005). 'The French Tableau de Bord and the American Balanced Scorecard: a critical analysis'. *Critical Perspectives on Accounting,* 16: 645–664.

Bevan, G. (2006). '*Setting Targets for Healthcare Performance: Lessons from Case Study of the English NHS'*. National Institute Economic Review, 197:67-79.

Bevan, G. and Hood, C. (2006). 'What's Measured is What Matters: Targets and Gaming in the English Public Health Care System'. *Public Administration,* 84(3):517-538.

Bhatt, N.V. (1996). 'Banks and income smoothing: an empirical analysis'. *Applied Financial Economics,* 6:505–510.

Bourn, M. and Ezzamel, M. (1986). 'Organisational Culture in Hospitals in the National Health Service'. *Financial Accountability and Management*, 2(3):203-225.

Braam, G. and Nijssen, E. (2004). 'Performance effects of using the balanced scorecard: the Dutch experience'. *Long Range Planning, 31:* 335-349.

Brickley, J.A. and Van Horn, R.L. (2002). 'Managerial Incentives in Non-profit Organizations: Evidence from Hospitals'. *Journal of Law and Economics,* 45: 225–49.

Brignall, S. and Modell, S. (2000). 'An institutional perspective on performance measurement and management in the 'new public sector''. *Management Accounting Research,* 11:281–306.

Bristol Enquiry (2001). '*The Report of the Public Inquiry into children's heart surgery at the Bristol Royal Infirmary 1984-1995: Learning from Bristol*'. CM 5207(I).

Broadbent, J. and Guthrie, J. (1992). 'Changes in the public sector: a review of recent 'alternative' accounting research'. *Accounting, Auditing and Accountability Journal*, 5(2): 3-31.

Broadbent, J. and Guthrie, J. (2008). 'Public Sector to Public Services: 20 years of "Contextual" Accounting Research'. *Accounting, Auditing and Accountability Journal*, 21 (2):129-169.

Broadbent, J. and Laughlin, R. (2002). 'Accounting choices: technical and political trade-offs and the UK's private finance initiative'. *Accounting, Auditing and Accountability Journal,* 15(5):622-654.

Broadbent, J. and Laughlin, R. (2005). 'Government concerns and tensions in accounting standard setting: the case of accounting for the Private Finance Initiative in the UK'. *Accounting and Business Research*, 35(3):207-228.

Broadbent, J. and Laughlin, R. (2009). 'Performance Management Systems: A Conceptual Model. *Management Accounting Research, 20*: 283-295.

Burgstahler, D.C. and Dichev, I. (1997). 'Earnings Management to Avoid Earnings Decreases and Losses'. *Journal of Accounting and Economics*, 24(1):99–126.

Canadian Institute for Health Information (2007). '*Hospital Report, 2007*, *Acute Care*'. ISBN 978-1-55465-059-0.

Chang, L-C. (2006). 'Managerial Responses to Externally Imposed Performance Measurement in the NHS: An Institutional Theory Perspective', *Financial Accountability & Management*, 22(1): 63–85.

Chang, L-C. (2007). 'The NHS performance assessment framework as a balanced scorecard approach: Limitations and implications, *International Journal of Public Sector Management*, 20(2): 101-117.

Chang, L-C. (2009). 'The Impact of Political Interests upon the Formulation of Performance Measurements: The NHS Star Rating System', *Financial Accountability & Management*, 25(2): 145-165.

Chirikos, T. N. (1998). 'Identifying efficiently and economically operated hospitals: The prospects and pitfalls of applying frontier regression techniques'. *Journal of Health Politics* 23 (6): 879-904.

Clatworthy, M. and Mellet, H. (1997). 'Managing Health and Finance: Conflict or Congruence?' *Public Money and Management,* October-December:41-46.

Clatworthy, M. A., Mellet, H. J. and Peel, M.J. (2002). 'The Market for External Audit Services in the Public Sector: An Empirical Analysis of NHS Trusts' *Journal of Business Finance & Accounting*, 29(9) & (10): 1399-1439.

Coase, R.H. (1937). '*The Nature of the Firm'*. Economica, 4:386-405.

Cohen, D.A., Dey, A. and Lys, T.Z. (2008). 'Real and Accrual-Based Earnings Management in the Pre- and Post-Sarbanes-Oxley Periods'. *The Accounting Review,* 83(3):757–787.

Collins, C., Kemsley D. and Lang, M. (1998). 'Cross Jurisdictional Income Shifting and Earnings Valuation. *Journal of Accounting Research*, 36(2):209-229.

Collins, J.H., Shackelford, D.A. and Wahlen J.M. (1995). 'Bank differences in the coordination of regulatory capital, earnings, and taxes'. *Journal of Accounting Research*, 33:263–291.

Combined Code (2003). '*Combined Code on Corporate Governance*' available at: http://www.fsa.gov.uk/pubs/ukla/lr_comcode2003.pdf

Cooper, Z., Gibbons, S., Jones, S. and McGuire, A. (2010). 'Does Hospital Competition Improve Efficiency? An Analysis of Recent Market-Based Reforms to the English NHS'. Centre for Economic Performance Discussion Paper No. 988, London School of Economics. ISSN 2042-2695.

Courty, P., Heinrich, C. and Marschke, G. (2005). 'Setting the standard in performance measurement systems'. *International Public Management Journal* 8(3): 321–347.

Courty, P. and Marschke, G. (2003a). 'Performance Funding in Federal Agencies: A Case Study of a Federal Job Training Program'. *Public Budgeting and Finance*, Fall 2003.

Courty, P. and Marschke, G. (2003b). 'Dynamics of Performance Measurement Systems'. *Oxford Review of Economic Policy*, 19(2):268-281.

Courty, P. and Marschke, G. (2004). 'An Empirical Investigation of Gaming Responses to Explicit Performance Incentives'. *Journal of Labor Economics,* 22(1):23-56.

Courty, P. and Marschke, G. (2007). 'Making Government Accountable: Lessons from a Federal Job Training Program'. *Public Administration Review*, Sep-Oct:904-916.

Courty, P. and Marschke, G. (2008). 'A general test for distortions in performance measures'. *The Review of Economics and Statistics,* 90(3): 428–441.

Davis, S. and Albright, T. (2004). 'An investigation of the effect of Balanced Scorecard implementation on financial performance'. *Management Accounting Research*, 15(2): 135-153.

Dechow, P.M., and Dichev, I. D. (2002). 'The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors'. *Accounting Review*, 77(Supplement):35–59.

Dechow, P. M., Richardson, S. and Tuna, I. (2001). 'Are Benchmark Beaters Doing Anything Wrong?' Working paper, University of Michigan.

Dechow, P. M., Richardson, S. and Tuna. I. (2003). 'Why Are Earnings Kinky? An Examination of the Earnings Management Explanation'. *Review of Accounting Studies*, 8(2/3):355–384.

Dechow, P.M. and Schrand, C.M. (2004). 'Earnings Quality'. Research Foundation of CFA Institute.

Dechow, P. M., and Sloan, R. G. (1991). 'Executive Incentives and the Horizon Problem: An Empirical Investigation'. *Journal of Accounting and Economics*, 14(1):51–89.

Dechow, P. M., Sloan, R. G. and Sweeney, A. P. (1995). 'Detecting Earnings Management'. *The Accounting Review*, 70(2):193–226.

Dechow, P. M., Sloan, R. G. and Sweeney, A. P. (1996). 'Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC'. *Contemporary Accounting Research,* 19: 1-36.

Degeorge, F., Patel, J. and Zeckhauser, R. (1999). 'Earnings management to exceed thresholds' *Journal of Business,* 72(1):1-33.

Deily, M.E. and McKay, N.L. (2006). 'Cost inefficiency and mortality rates in Florida hospitals'. *Health Economics*, 15: 419–431.

Demski, J. (1998). 'Performance Measure Manipulation'. *Contemporary Accounting Research*, 15:261-285.

Demski, J., Frimor, H. and Sappington, D. (2004). **'**Efficient Manipulation in a Repeated Setting'. *Journal of Accounting Research*, 42 (1): 31-49.

Demski, J. and Sappington, D.E.M. (1987). 'Delegated expertise'. *Journal of Accounting Research*, 25(1):68-89.

Demski, J. and Sappington, D.E.M, (1990). 'Fully revealing income measurement'. *The Accounting Review*, 65(2):363-383.

Department of Health, (1997). *'The New NHS'* Cm3807.

Department of Health, (1999). '*The NHS Performance Assessment Framework'*. 16431.

Department of Health, (2000). '*The NHS Plan'* Cm 4818-I.

Department of Health, (2002a). '*Delivering the NHS Plan'*, Cm 5503.

Department of Health, (2002b). '*Reforming NHS Financial Flows: Introducing Payment by Results*'. 29600.

Department of Health, (2002c). '*Raising standards across the NHS: a programme of rewards and support for all NHS Trusts*'. Ref: 30131.

Department of Health, (2003a). '*Payment by Results Consultation'*, Gateway Ref: 1632.

Department of Health, (2003b). '*Building on the best: choice, responsiveness and equity in the NHS*'. HMSO Cm 6079.

Department of Health, (2003c). '*Raising Standards, Improving Performance in the NHS*' 31829.

Department of Health, (2005). '*Creating a patient led NHS: Delivering the NHS Improvement Plan*', Gateway Reference: 4699.

Department of Health, (2006a). '*Reference Costs 2005-06'* available at: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_118338.pdf

Department of Health, (2006b). '*Reference Costs 2004-05'* available at:

http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4133221

Department of Health, (2007). *The Government's Response to the Health Select Committee's Report on NHS Deficits*. The Stationery Office, Cm 7028.

Department of Health, (2010). '*Review of Reference Costs – Summary Repor*t' Gateway ref: 14618.

De Vries, J. (2010). 'Is New Public Management Really Dead'. *OECD Journal on Budgeting*, 1:87-91.

Dunleavy, P., Margetts, H., Bastow, S. and Tinkler, J. (2005). 'New Public Management Is Dead—Long Live Digital-Era Governance'. *Journal of Public Administration Research and Theory,* 16:467–494.

Durtschi, C. and Easton, P. (2005). 'Earnings Management? The Shapes of the Frequency Distributions of Earnings Metrics Are Not Evidence *ipso facto'*. *Journal of Accounting Research,* 43(4): 557–92.

Durtschi, C. and Easton, P. (2009). 'Earnings Management? Erroneous Inferences Based on Earnings Frequency Distributions'. *Journal of Accounting Research,* 47(5): 1249-1281.

Edwards, P. and Shaoul, J. (2003). 'Partnerships: for better, for worse?' Accounting, Auditing and Accountability Journal, 16(3):397-421.

Eldenburg, L., Gunny, K., Hee, K. and Soderstrom, N. (2007). 'Earnings Management through Real Activities Manipulation: Evidence from Non-profit Hospitals', Working paper.

Eldenburg, L., Hermalin, B.E., Weisbach, M. S. and Wosinska, M. (2004). 'Governance, performance objectives and organizational form: evidence from hospitals'. *Journal of Corporate Finance*, 10(4):527– 548.

Eldenburg, L. and Kallapur, S. (1997). 'Changes in hospital service mix and cost allocations in response to changes in Medicare reimbursement schemes'. *Journal of Accounting and Economics*, 23:31- 51.

Eldenburg, L. and Krishnan, R. (2003). 'Public versus private governance: a study of incentives and operational performance'. *Journal of Accounting and Economics*, 35:377–404.

Eldenburg, L. and Krishnan, R. (2008). 'The influence of ownership on accounting information expenditures'. *Contemporary Accounting Research*, 25(3):739-772.

Eldenburg, L. and Vines, C.C. (2004). 'Nonprofit classification decisions in response to a change in accounting rules'. *Journal of Accounting and Public Policy*, 23:1–22.

Ellwood, S. (1995). 'NHS Costing for Contracting Rules', *Public Money and Management*, April-June 1995.

Ellwood, S. (1996a). 'Pricing Services in the UK National Health Service', *Financial Accountability and Management*, 12(4):281-301.

Ellwood, S. (1996b). 'Full-cost pricing rules within the National Health Service internal market—accounting choices and the achievement of productive efficiency' *Management Accounting Research,* 7:25-51.

Ellwood, S. (2003). 'Bridging the GAAP across the UK public sector'. *Accounting and Business Research*, 33 (2): 105-121.

Ellwood. S. (2008). 'Accounting for Public Hospitals: A Case Study of Modified GAAP'. *Abacus,* 44(4):399-422.

Ellwood, S. and Newbury, S. (2006). 'A bridge too far: a common conceptual framework for commercial and public benefit entities'. *Accounting and Business Research*, 36(1):19-32.

Emmanuel, C. (1999). 'Income Shifting and International Transfer Pricing: A Three-Country Example.' *Abacus*, 35(3):252-267.

Enthoven, A. (1985). '*Reflections on the management of the National Health Service: an American looks at incentives to efficiency in health services management in the UK'*. Nuffield Provincial Hospitals Trust, London.

Ezzamel, M. and Willmot, H. (1993). 'Corporate Governance and Accountability: Recent Reforms in the UK Public Sector'. Accounting*, Auditing and Accountability Journal*, 6 (3): 109-132.

Feltham, G.A. and Xie, J. (1994). 'Performance Measure Congruity and Diversity in Multi-task Principal/Agent Relations'. *The Accounting Review*, 69(3):429-453.

Ferrari, A. (2006). 'Market oriented reforms of health services: a non-parametric analysis'. *The Service Industries Journal.* 26(1):1-13.

Financial Reporting Council (2006). '*The Combined Code on Corporate Governance'*. ISBN: 9 781841 40816.

Folland, S.T. and Hofler, R.A. (2001). 'How reliable are hospital efficiency estimates? Exploiting the dual to homothetic production'. *Health Economics*, 10:683-698.

Forgione, D.A. (1997). 'Health Care Financial and Quality Measures: International Call for a 'Balanced Scorecard Approach'. *Journal of Health Care Finance*, 24(1):55-58.

Foster, G. (1979). 'Briloff and the capital market'. *Journal of Accounting Research, 17*: 262-274.

Gaver, J. J., Gaver, K. M. and Austin, J. R. (1995). 'Additional Evidence on Bonus Plans and Income Management'. *Journal of Accounting and Economics,* 19(1):3–28.

Giuffrida, A. and Gravelle, H. (2001). 'Measuring performance in primary care: Econometric analysis and DEA'. *Applied Economics, 33* (2): 163-75.

Goddard, A. (2010). ‘Contemporary public sector accounting research – An international comparison of journal papers’. *British Accounting Review*, 42:75-87.

Griffiths, R. (1983). *NHS Management Inquiry, (The Griffiths Report),* London: Department of Health and Social Security. Available at: http://www.nhshistory.com/griffiths.html

Groene, O., Klazinga, N., Kanzandjian, V., Lombrail, P., Bartels, P. (2008). The World Health Organization Performance Assessment Tool for Quality Improvement in Hospitals (PATH): An Analysis of the Pilot Implementation in 37 Hospitals. *International Journal for Quality in Health Care, 20*(3): 155-161.

Groot, T. and Budding, T. (2008), ‘New Public Management’s Current Issues and Future Prospects’. *Financial Accountability & Management*, 24(1): 1-13.

Grubert, H. (2003). ‘Intangible Income, Intercompany Transactions, Income Shifting, and the Choice of Location’. *National Tax Journal,* 56(1):221-242.

Guven-Uslu, P. and Conrad, L. (2008). ‘Uses of Management Accounting Information for Benchmarking in NHS Trusts’. *Public Money and Management,* 28(4):239-246.

Hall, M. (2008). 'The effect of comprehensive performance measurement systems on role clarity, psychological empowerment and managerial performance'. *Accounting*, *Organizations and Society,* 33 (2/3): 141-163.

Hauck, K. and Street, A. (2006). ‘Do targets matter? A comparison of English and Welsh national health priorities’. *Health Economics,* 16: 275–290.

Hayn, C. (1995). ‘The Information Content of Losses’. *Journal of Accounting and Economics,* 20(2):125–153.

Healy, P. M. (1985). ‘The Effect of Bonus Schemes on Accounting Decisions’. *Journal of Accounting and Economics,* 7 (1–3):85–107.

Healy, P. M. and. Wahlen, J. M. (1999). 'A Review of the Earnings Management Literature and Its Implications for Standard Setting'. *Accounting Horizons,* 13(4):365–383.

Healthcare Commission, (2005a). '*Assessment for Improvement: The Annual Health Check*'. ISBN 1-84-562017-8.

Healthcare Commission (2005b). '*NHS Performance ratings: 2004/2005*'. ISBN 1-84562-049-6.

HM Treasury, Financial Reporting Manual available at: http://www.hmtreasury.gov.uk/frem_previous_manuals.htm

Hodges, R., MacNiven, L. and Mellett, H. (2004). *'*Governance of UK NHS Trusts: The Annual General Meeting'. *Corporate Governance*, 12(3):343-352.

Hodges, R. and Mellett, H. (2002).' Investigating standard setting: accounting for the United Kingdom's private finance initiative'. *Accounting Forum*, 26(2): 126-151.

Hodges, R. and Mellett, H. (2004). 'Reporting PFI in Annual Accounts: A User's Perspective'. *Public Money and Management*, June: 153-158.

Hodges, R. and Mellett, H. (2005). 'Accounting for the U.K.'s Private Finance Initiative: An Interview-Based Investigation'. *Abacus*, 41(2):159-180.

Hoerger, T.J. (1991). 'Profit' variability in for-profit and not-for-profit hospitals'. *Journal of Health Economics,* 10:259–289.

Hollingsworth, B. (2008). 'The measurement of efficiency and productivity of health care delivery'. *Health Economics*, 17 (10):1107-1128.

Holmstrom, B. (1979). 'Moral Hazard and Observability'. *The Bell Journal of Economics*, 10 (1):74-91.

Holthausen, R. W., Larcker, D. F. and Sloan, R. G. (1995). 'Annual Bonus Schemes and the Manipulation of Earnings'. *Journal of Accounting and Economics*, 19(1):29–74.

Hood, C. (1991). 'A Public Management for All Seasons'. *Public Administration*, 69:3-19.

Hood, C. (1995). 'The "New Public Management" in the 1980s: Variations on a Theme'. *Accounting, Organizations and Society*, 20 (2/3):93-109.

Hopwood, A. (1972), 'An empirical study of the role of accounting information in performance evaluation', *Journal of Accounting Research*, Supplement, 156-182.

Hoque, Z. and James, W. (2000). 'Linking Balanced Scorecard Measures to Size and Market Factors: Impact on Organizational Performance'. *Journal of Management Accounting Research,* 12:1-17.

House of Commons (2006a). *NHS Deficits, First report of Session 2006-07: Volume 1*. House of Commons Health Committee, The Stationery Office.

House of Commons (2006b). *NHS Deficits, First report of Session 2006-07: Volume II Oral and written evidence.* House of Commons Health Committee. The Stationery Office.

Iliev, P. (2010). 'The Effect of SOX Section 404: Costs, Earnings Quality, and Stock Prices'. *The Journal of Finance*, 65(3):1163-1196.

Ittner, C.D. (2008). 'Does measuring intangibles for management purposes improve performance? A review of the evidence'. *Accounting and Business Research*, 38(3):261-272.

Ittner, C.D. and Larcker, D.F. (1998a). 'Innovations in Performance Measurement: Trends and Research Implications', *Journal of Management Accounting Research*, 10:205–38.

Ittner, C.D. and Larcker, D.F. (1998b). 'Are Nonfinancial Measures Leading Indicators of Financial Performance? An Analysis of Customer Satisfaction'. *Journal of Accounting Research*, 36: 1-35.

Ittner, C.D. and Larcker, D.F. (2003). 'Coming up short on nonfinancial measurement'. *Harvard Business Review*, November: 88-95.

Ittner, C.D., Larcker, D.F. and Randall, T. (2003). 'Performance implications of strategic performance measurement in financial services firms'. *Accounting Organizations and Society*, 28:715-741.

Jacobs, R. (2001). 'Alternative Methods to Examine Hospital Efficiency: Data Envelopment Analysis and Stochastic Frontier Analysis'. *Health Care Management Science* 4:103–115.

Jain, P.K. and Rezaee, Z. (2006). 'The Sarbanes-Oxley Act of 2002 and Capital-Market Behaviour: Early Evidence'. *Contemporary Accounting Research*, 23(3):629-654.

Jensen, M. and Meckling, W.H. (1976). 'Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure'. *Journal of Financial Economics,* 3 (4): 305-360.

Jones, C.S. (2002). 'The attitudes of British National Health Service managers and clinicians towards the introduction of benchmarking'. *Financial Accountability and Management*, 18(2):163-188.

Jones, C.S. and Dewing, I.P. (1997). 'The attitudes of NHS clinicians and medical managers towards changes in accounting controls'. *Financial Accountability and Management*, 13(3):261-280.

Jones, J. F. (1991). 'Earnings Management during Import Relief Investigations'. *Journal of Accounting Research*, 29(2):193–228.

Jones, M.L. and Filip, S.J. (2000). 'Implementation and Outcomes of a Balanced Scorecard Model in Women's Services in an Academic Health Care Institution'. *Quality Management in Health Care*, 8(4):40-51.

Jones, M.J. and Mellett, H.J. (2007). 'Determinants of changes in accounting practices: Accounting and the UK Health Service' *Critical Perspectives on Accounting*, 18:91–121.

Kallapur, S. and Eldenburg, L. (2005). 'Uncertainty, Real Options, and Cost Behavior: Evidence from Washington State Hospitals'. *Journal of Accounting Research*, 43(5):735-752.

Kanji, G.K. (1998). 'Measurement of Business Excellence', *Total Quality Management*, 9: 633-643.

Kaplan, R. S. (2001). 'Strategic Performance Measurement and Management in Nonprofit Organisations'. *Nonprofit Management and Leadership,* Spring: 353-370.

Kaplan, R. and Norton, D. (1992). 'The Balanced Scorecard – Measures That Drive Performance'. *Harvard Business Review,* 70(1):71-79.

Kaplan, R. and Norton, D. (1996a). 'Using the Balanced Scorecard as a Strategic Management System'. *Harvard Business Review,* 74(1):75-85.

Kaplan, R. S. and Norton, D. P. (1996b). *The Balanced Scorecard—translating strategy into action'*. Boston: Harvard Business School Press.

Kaplan, R. and Norton, D. (2000). 'Having trouble with your strategy? Then map it'. *Harvard Business Review,* 78(5):167-176.

Kasznik, R. (1999). 'On the Association between Voluntary Disclosure and Earnings Management'. *Journal of Accounting Research*, 37(1):57–82.

Kelman, S. and Friedman, J.N. (2009). 'Performance Improvement and Performance Dysfunction: An Empirical Investigation of Distortionary Impacts of the Emergency Room Wait-Time Target in the English National Health Service'. *Journal of Public Administration Research and Theory*, 19:917-946.

Key, K. G. (1997). 'Political Cost Incentives for Earnings Management in the Cable Television Industry'. *Journal of Accounting and Economics,* 23(3):309–338.

Klassen, K., Lang, M. and Wolfson, M. (1993). 'Geographic Income Shifting by Multinational Corporations in Response to Tax Rate Changes'. *Journal of Accounting Research*, 31(Supplement):141-173.

Klein, R. (2006). '*The New Politics of the NHS'*, 5th Edition Radcliffe Publishing ISBN 978-1846 190667.

Klein, R. (2010). '*The Eternal Triangle: Sixty Years of the Centre–Periphery Relationship in the National Health Service*'. Social Policy and Administration, 44(3):285-304.

Kloot, L. (2009). 'Performance measurement and accountability in an Australian fire service'. *International Journal of Public Sector Management*, 22(2):128-145.

Lapsley, I. (1992). 'Public Services in Transition: the Introduction of Quasi-markets'. *Financial Accountability and Management*, 8(2):75-76.

Lapsley, I. (1993). 'Markets Hierarchies and the Regulation of the National Health Service'. *Accounting and Business Research,* 23 (91A):384-394.

Lapsley, I. (1994). 'Market mechanisms and the Management of Health Care'. *International Journal of Public Sector Management, 7(6):15-25.*

Lapsley, I. (2008). 'The NPM Agenda: Back to the Future'. *Financial Accountability and Management*, 24 (1):77-96.

Le Grand, J. (1999). 'Competition, Cooperation, Or Control? Tales From the British National Health Service'*. Health Affairs*, May/June: 27-39.

Le Grand, J. (2002a). 'The Labour Government and the National Health Service'. *Oxford Review of Economic Policy*, 18(2):137-153

Le Grand, J. (2002b). 'Further tales from the British Health Service'*. Health Affairs*, 21(3):116-128.

Le Grand, J. (2003). '*Motivation, agency, and public policy: of knights and knaves, pawns and queens'*. Oxford University Press, ISBN 0199266999.

Le Grand, J. (2007). '*The Other Invisible Hand: Delivering Public Services Through Choice and Competition'*. Oxford: Princeton University Press.

Le Grand, J. (2010). '*Knights and Knaves Return: Public Service Motivation and the Delivery of Services'*. International Public Management Journal*,* 13(1): 56–71.

Leone, A.J. and Van Horn, R.L. (2005). 'How do non-profit hospitals manage earnings?' *Journal of Health Economics,* 24:815–837.

Likierman, A. (1998a). 'Resource Accounting and Budgeting – Where are we now?' *Public Money and Management*, April-June: 17-20.

Likierman, A. (1998b). 'Report – Recent Developments in Resource Accounting and Budgeting'. *Public Money and Management,* October- December: 62-64.

Li, H., Pincus, M. and Rego, S.O. (2008). 'Market Reaction to Events Surrounding the Sarbanes-Oxley Act of 2002 and Earnings Management'. *Journal of Law and Economics*, 51:111-134.

Llewellyn, S. and Northcott, D. (2005). 'The average hospital'. *Accounting, Organizations and Society*, 30:555-583.

Lobo G.J., and Yang, D-H. (2001). 'Bank managers' heterogeneous decisions on discretionary loan loss provisions'. *Review of Quantitative Finance and Accounting,* 16:223–250.

Ludwig, M., Van Merode, F. and Groot, M. (2010). 'Principal agent relationships and the efficiency of hospitals'. *European Journal of Health Economics*, 11:291–304.

Lynch, R.L. and Cross, K.F. (1991). '*Measure up! Yardsticks for Continuous Improvement*', Basil Blackwell Inc., Cambridge, MA.

Makar, S. D., Alam, P. and Pearson, M. A. (1998). 'Antitrust Merger Investigations and the Quality of Reported Earnings'. *Journal of Applied Business Research*, 14(4):89–100.

Malina, M.A., Nørreklit, H.S.O. and Selto, F.H. (2008). 'Relations among Measures, Climate of Control, and Performance Measurement Models'. *Contemporary Accounting Research,* 24(3):935-982.

Maniadakis, N., Hollingsworth, B. and Thanassoulis, E. (1999). 'The impact of the internal market on hospital efficiency, productivity and service quality'. *Health Care Management Science,* 2:75–85.

Maniadakis, N. and Thanassoulis, E. (2000). 'Assessing productivity changes in UK hospitals reflecting technology and input prices'. *Applied Economics*, 32:1575-1589.

Mayston, D.J. (1985). ‘Non-profit Performance Indicators in the Public Sector’. *Financial Accountability and Management*, 1(1):51-74

McCallion, G., McKillop, D. G., Glass, J.C. and Kerr, C. (1999). ‘Rationalising Northern Ireland hospital services towards larger providers: best practice efficiency studies and current policy’. *Public Money and Management,* 19:27-32.

McCallion, G., Glass, J.C., Jackson, R., Kerr, C. and McKillop, D.G. (2000). ‘Investigating productivity change and hospital size: a non-parametric frontier approach’. *Applied Economics,* 32:161-174.

McNichols, M.F. (2000). ‘Research Design Issues in Earnings Management Studies’, *Journal of Accounting and Public Policy*, 19(4–5):313–45.

McNichols, M. F. and Stubben, S.R. (2008). ‘Does Earnings Management Affect Firms' Investment Decisions?’ *The Accounting Review,* 83(6):1571-1603.

Meliones, J.N. (2000). ‘Saving Money, Saving Lives’. *Harvard Business Review,* 78(6):57-62, 64, 66-67.

Mohan, J. (2002). *‘Planning, Markets and Hospitals’*. Routledge. London. ISBN 0-415-19607-8.

Modell, S. (2009). ‘Institutional research on performance measurement and management in the public sector accounting literature: a review and assessment’ *Financial Accountability and Management*, 25(3):277-303.

Mutter, R.L., Rosko, M.D., Greene, W.H. and Wilson, P.W. (2011). ‘Translating Frontiers Into Practice: Taking the Next Steps Toward Improving Hospital Efficiency’. *Medical Care Research and Review*, 68(1) Supplement: 3S–19S.

National Audit Office (2004). ‘*NHS (England) Summarised Accounts 2002-03*, HC 505-I.

National Health Service and Community Care Act (1990) available at: http://www.opsi.gov.uk/ACTS/acts1990/Ukpga_19900019_en_1.htm

Neely, A., Adams, C. and Kennerley, K. (2002). '*The Performance Prism: The Scorecard For Measuring And Managing Business Success'*, Financial Times Prentice Hall, London.

Nelson,M., Elliott, J. and Tarpley, R. (2003). 'How Are Earnings Managed? Examples from Auditors." *Accounting Horizons,* 17:17–35.

Newberry, K.J. and Dhaliwal, D.S. (2001). 'Cross-Jurisdictional Income Shifting by U.S. Multinationals: Evidence from International Bond Offerings.' *Journal of Accounting Research,* 39(3):643-662.

Newberry, S., and Pallot, J. (2003). 'Fiscal (ir)responsibility: privileging PPPs in New Zealand'. *Accounting, Auditing and Accountability Journal,* 16 (3):467-492.

NHS Executive (1999). '*Guidance to Health Authorities and NHS Trusts on Break-even duty, Provisions and Accumulated Deficits'*. HSC 1999/146.

Nørreklit, H. (2000). **'**The Balance on the Balanced Scorecard— A Critical Analysis of Some of its Assumptions'. *Management Accounting Research*, 11(1): 65-88.

Nørreklit, H. (2003). 'The Balanced Scorecard: what is the score? A rhetorical analysis of the balanced scorecard'. *Accounting, Organizations and Society*, 28:591-619.

Nørreklit, H., Jacobsen, M. and Mitchell, F. (2008), 'Pitfalls in using the balanced scorecard'. *Journal of Corporate Accounting and Finance,* 19(6):65-68.

Northcott, D. and Llewellyn, S. (2003). 'The 'ladder of success' in healthcare: the UK national reference costing index'. *Management Accounting Research*, 14:51-66.

Northcott, D. and Llewellyn, S. (2005). *'*Benchmarking in UK health: a gap between policy and practice?*' Benchmarking: An International Journal*, 12(5): 419-435.

O'Neill, L., Rauner, M., Heidenberger, K. and Kraus, M. (2008). 'A cross-national comparison and taxonomy of DEA-based hospital efficiency studies'. *Socio-Economic Planning Sciences,* 42:158–189.

Otley, D. (2008). 'Did Kaplan and Johnson get it right?' *Accounting, Auditing and Accountability Journal*, 21(2):229-329.

Ouchi, W.G. (1979). 'A conceptual framework for Design of Organisational Control Mechanisms'. *Management Science*, 25(9):833-848.

Ouchi, W.G. (1980). 'Markets, Bureaucracies and Clans'. *Administrative Science Quarterly*, 25:129-141.

Oyelere, P.B. and Emmanuel, C. R. (1998). 'International transfer pricing and income shifting: evidence from the UK'. *The European Accounting Review*, 7(4): 623-635.

Palmrose, Z-V., Richardson, V. and Scholz. S. (2004). 'Determinants of market reactions to restatement announcements'. *Journal of Accounting and Economics,* 37: 59-89.

Park, C. (1958). 'Thought processes in creative accounting'. *The Accounting Review,* 441-445*.*

Parkin, D. and Hollingsworth, B. (1997). 'Measuring production efficiency of acute hospitals in Scotland, 1991-1994: validity issues in data envelopment analysis'. *Applied Economics*, 29:1425-1433.

Phelps, M.G. (2009). 'Changing costs of public services'*. Economic & Labour Market Review*, 3(12).

Pink, G.H., McKillop, I., Schraa, E.G., Preyra, C., Montgomery, C. and Baker, G.R. (2001). 'Creating a Balanced Scorecard for a Hospital System'. *Journal of Health Care Finance*, 27(3):1-20.

Pollitt, C. (1986). 'Beyond the managerial model: the case for broadening performance assessment in Government and the public services'. *Financial Accountability and Management,* 2(3):155-170.

Pollitt, C., Harrison, S. Hunter, D.J. and Marnoch, G. (1991). General management in the NHS: the initial impact 1983-1988. *Public Administration,* 69:61-83.

Pollock, A. (2004). *'NHS Plc: The privatisation of our health care'*, London, Verso. ISBN 1844670112.

Preston, A.M., Cooper, D.J., and Coombs, R.W. (1992). 'Fabricating budgets: a study of the production of management budgeting in the National Health Service'. *Accounting, Organizations and Society*, 17(6): 561-593.

Propper, C., Sutton, M., Whitnall, C. and Windjmeier, F. (2010). 'Incentives and targets in hospital care: Evidence from a natural experiment'. *Journal of Public Economics*, 94:318-335.

Propper, C. and Wilson, D. (2003). 'The use and usefulness of performance measures in the public sector' *Oxford Review of Economic Policy*, 19(2):250-267.

Radnor, Z. and Lovell, B. (2003a). 'Success factors for implementation of the balanced scorecard in a NHS multi-agency setting'. *International Journal of Health Care Quality Assurance,* 16(2):99-108.

Radnor, Z. and Lovell, B. (2003b). 'Defining, justifying and implementing the Balanced Scorecard in the National Health Service'. *International Journal of Medical Marketing,* 3(3):174–188.

Rivett, G.C. (1998). '*From Cradle to Grave: fifty years of the NHS'*. London, King's Fund. ISBN 1-85717-148-9.

Robson, N. (2003). 'From voluntary to state control and the emergence of the department in UK hospital accounting'. *Accounting, Business and Financial History,* 13(2):99-123.

Rosko M. D. and Mutter R. L. (2011). 'What Have We Learned From the Application of Stochastic Frontier Analysis to U.S. Hospitals?' *Medical Care Research and Review*, 68(1):75S-100S.

Rutherford, B. (2003). 'The Social Construction of Financial Statement Elements under Private Finance Initiative Schemes'. *Accounting, Auditing and Accountability Journal,* 16(3):372-396.

Said, A. A., HassabElnaby, H. R. and Wier, B. (2003). 'An Empirical Investigation of the Performance Consequences of Nonfinancial Measures'. *Journal of Management Accounting Research*, *15*, 193–223.

Schipper, K. (1989). 'Earnings Management'. *Accounting Horizons*, 3(4):91–102.

Schwartz, M., Cohen, A. B., Restuccia, J. D., Ren, Z. J., Labonte A., Theokary, C., Kang, R. and Horwitt, J. (2011). 'How Well Can We Identify the High-Performing Hospital?' *Medical Care Research and Review*, 68(3) 290–310.

Scott, W.R. (2009). *Financial Accounting Theory*, 5th Edition. Pearson Prentice Hall, ISBN 13:978-0-13-207286-1.

Smith, P. (1995). 'On the Unintended Consequences of Publishing Performance Data in the Public Sector'. *International Journal of Public Administration*, 18:277-310.

Smith, R.E. & Wright, W.F. (2004). 'Determinants of Customer Loyalty and Financial Performance'. *Journal of Management Accounting Research*, 16:183-205.

Stocken, P.C. and Verrecchia, R.E. (2004). 'Financial Reporting System Choice and Disclosure Management'. *The Accounting Review,* 79(4):1181-1203.

Street, A. (2003). 'How much confidence should we place in efficiency estimates?' *Health Economics,* 12 (11): 895-907.

Street, A. and Jacobs, R. (2002). 'Relative performance evaluation of the English acute hospital sector'. *Applied Economics,* 34 (9): 1109-1119.

Stubben, S.R. (2010). 'Discretionary Revenues as a Measure of Earnings Management'. *The Accounting Review,* 85(2): 695–717.

Talbot-Smith, A. and Pollock, A. M. (2006). '*The New NHS – a guide'*. London, Routledge. ISBN 0-415-32841-0.

Teoh, S., Wong, T. and Rao, G. (1998). 'Are accruals during initial public offerings opportunistic?' *Review of Accounting Studies,* 3: 175-208.

Thanassoulis, E., Boussofiane, A. and Dyson, R. G. (1996). 'A comparison of data envelopment analysis and ratio analysis as tools for performance measurement'. *OMEGA, International Journal of Management Science,* 24(3): 229-44.

Thomas, W.B., Herrmann, D.R. and Inoue, T. (2004). 'Earnings management through affiliated transactions'. *Journal of International Accounting Research*, 3(2):1-25.

Van Helden, G. J. (2005). 'Researching public sector transformation: The role of management accounting'. *Financial Accountability & Management,* 21(1):99-133.

Vinnari, E.M. and Nasi, S. (2008). 'Creative accrual accounting in the public sector: milking water utilities to balance municipal budgets and accounts'. *Financial Accountability and Management,* 24(2):97-116.

Wahlen J. M. (1994). 'The nature of information in commercial bank loan loss disclosures'. *The Accounting Review,* 69:455–478.

Walker, R.G. and Robinson, P. (1993). 'A Critical Assessment of the Literature on Political Activity and Accounting Regulation'. *Research in Accounting Regulation,* 7:3-40.

Watkin, B. (1978). '*The National Health Service: The first phase'*. George Allen and Unwin. ISBN 0-04-362025-6.

Webster, C. (2002). '*The National Health Service – A political history'*. Oxford University Press. ISBN 9 7801990251100.

Williams, P., Gregory-Jenkins, J., and Ingraham, L. (2006). 'The winnowing away of behavioral accounting research in the US: the process for anointing academic elites'. *Accounting, Organizations and Society,* 31(8): 783–818.

Williamson, O. E. (1975), *Markets and Hierarchies,* Free Press.

World Health Organization (2003a). '*How can hospital performance be measured and monitored?'* WHO Regional Office for Europe's Health Evidence Network (HEN).

World Health Organization (2003b). '*Measuring hospital performance to improve the quality of care in Europe: a need for clarifying the concepts and defining the main dimensions'*. Report on a WHO Workshop, Barcelona, Spain, 10-11 January 2003.

Worthington, A. C. (2004). 'Frontier Efficiency Measurement in Health Care: A Review of Empirical Techniques and Selected Applications'. *Medical Care Research and Review***,** 61(2*)*:135-170.

Zelman, W.N., Pink, G.H. and Matthias, C.B. (2003). 'Use of the balanced scorecard in healthcare'. *Journal of Healthcare Finance*, 29(4):1-16.

CHAPTER 9 APPENDICES

APPENDIX 1 PERFORMANCE MEASURES IN THE STAR RATINGS SYSTEM 2003-2005

PERFORMANCE MEASURES IN THE STAR RATINGS SYSTEM 2003-2005		
2002-03	**2003-04**	**2004-05**
Key measures	**Key measures**	**Key measures**
12 hour waits for emergency admission via A&E post decision to admit	12 hour waits for emergency admission via A&E post decision to admit	12 hour waits for emergency admission via A&E post decision to admit
Hospital cleanliness	Hospital cleanliness	Hospital cleanliness
Improving Working Lives	Improving Working Lives	Improving Working Lives[54]
Outpatients waiting longer than the standard	Outpatients waiting longer than the standard	Outpatients waiting longer than the standard
Total time in A&E: 4 hours or less	Total time in A&E: 4 hours or less	Total time in A&E: 4 hours or less
Patients waiting longer than the standard for elective admission (inpatients)	Patients waiting longer than the standard for elective admission (inpatients)	Patients waiting longer than the standard for elective admission (inpatients)
All Cancers: 2 week wait	All Cancers: 2 week wait	All Cancers: 2 week wait
Cancelled operations not admitted within 28 days	Outpatient and elective (inpatient and day case) booking	Outpatient and elective (inpatient and day case) booking
Financial management	Financial management	Financial management
Balanced scorecard measures	**Balanced scorecard measures**	**Balanced scorecard measures**
Patient focus	***Patient focus***	***Patient focus***
A&E emergency admission waits (4 hours)	A&E emergency admission waits (4 hours)	A&E emergency admission waits (4 hours)
Better hospital food	Better Hospital Food	Better Hospital Food
Breast cancer: 1 month diagnosis to treatment	Breast cancer: 1 month diagnosis to treatment	Breast cancer: 1 month diagnosis to treatment
Delayed transfers of care	Delayed transfers of care	
Cancelled operations	Cancelled operations	Cancelled operations
Day case patient booking	Day case patient booking	
Thirteen week outpatient waits	Thirteen week outpatients	

[54] The indicator for Improving Working Lives was originally retained for 2004-05 but was subsequently dropped as the processes associated with this indicator were fully implemented across all Trusts by 2003-04.

PERFORMANCE MEASURES IN THE STAR RATINGS SYSTEM 2003-2005		
2002-03	**2003-04**	**2004-05**
Patient focus – cont'd		
Six month inpatient waits	Six month inpatient waits	
Thirteen week outpatient waits	Thirteen week outpatients	
Patient complaints	Patient complaints	
Nine month heart operation waits	Breast cancer: 2 month GP urgent referral to treatment	Breast cancer: 2 month GP urgent referral to treatment
Outpatient A&E survey - access & waiting	Adult inpatient and young patient surveys: access and waiting	Adult inpatient and young patient surveys: access and waiting
Outpatient A&E survey - better information, more choice	Adult inpatient and young patient surveys: better information, more choice	Adult inpatient and young patient surveys: better information, more choice
Outpatient A&E survey - building relationships	Adult inpatient and young patient surveys: building closer relationships	Adult inpatient and young patient surveys: building closer relationships
Outpatient A&E survey - clean, comfortable, friendly place to be	Adult inpatient and young patient surveys: clean, comfortable, friendly place to be	Adult inpatient and young patient surveys: clean, comfortable, friendly place to be
Outpatient A&E survey - safe, high quality, co-ordinated care	Adult inpatient and young patient surveys: safe, high-quality, coordinated care	Adult inpatient and young patient surveys: safe, high-quality, coordinated care
Paediatric outpatient did not attend rates	Patients waiting longer than standard for revascularisation	
Privacy & dignity		
Total inpatient waits		
Waiting times for Rapid Access Chest Pain Clinic		
Clinical focus	***Clinical focus***	***Clinical focus***
Clinical Negligence	Clinical negligence	Clinical negligence
Deaths following a heart bypass operation	Deaths following a heart bypass operation	Deaths following a heart bypass operation
Deaths following selected non-elective surgical procedures	Deaths following selected non-elective surgical procedures	Deaths following selected non-elective surgical procedures
Emergency readmission following discharge (adults)	Emergency readmission following discharge (adults)	Emergency readmission following discharge (adults)
Emergency readmission to hospital following treatment for a fractured hip	Emergency readmission following discharge for a fractured hip	Emergency readmission following discharge for a fractured hip

PERFORMANCE MEASURES IN THE STAR RATINGS SYSTEM 2003-2005		
2002-03	**2003-04**	**2004-05**
Clinical focus – cont'd		
	Child protection	Child protection
	Clinical governance composite indicator	Clinical governance composite indicator
Thrombolysis - 30 minute door to needle time	Thrombolysis - 30 minute door to needle time	Thrombolysis - 30 minute door to needle time
	Composite of participation in audits	Composite of participation in audits
Emergency readmission to hospital following discharge for children		
Emergency readmission to hospital following treatment for a stroke	Indicator on stroke care	Indicator on stroke care
Infection control procedures		
Methicillin Resistant *Staphylococcus Aureus* (MRSA) bacteraemia improvement score	Infection control	Infection control
	"Winning Ways" - processes and procedures	"Winning Ways" - processes and procedures
Capacity and capability	***Capacity and capability***	***Capacity and capability***
Consultant appraisal	Consultant appraisal	Consultant appraisal
Information Governance	Information governance	Information governance
Data quality	HES & Workforce datasets: data quality on ethnic group	HES & Workforce datasets: data quality on ethnic group
Junior doctors' hours	Junior doctors' hours	Junior doctors' hours
Fire, Health & Safety	Staff opinion survey: health, safety and incidents	Staff opinion survey: health, safety and incidents
Sickness absence rate	Staff opinion survey: human resource management	Staff opinion survey: human resource management
Staff opinion survey	Staff opinion survey: staff attitudes	Staff opinion survey: staff attitudes

APPENDIX 2 SCORING METHODOLOGY FOR THE KEY TARGETS AND THE BALANCED SCORECARD IN THE STAR RATINGS SYSTEM

Key Targets

A trust's performance on each Key Target is assessed as either 'achieved', 'underachieved' or 'significantly underachieved'. This is based on expected levels of performance as defined in government policy using two defined thresholds. The first threshold distinguishes between achieved and underachieved; the second distinguishes between underachieved and significantly underachieved.

For each Key Target, a trust is allocated penalty points in relation to its performance level using the following rules:

Achieved target:	0 points
Underachieved:	2 points
Significantly underachieved:	6 points

Thus, significantly underachieving on one Key Target is equivalent to underachieving on three Key Targets, reflecting the seriousness of failure to meet the desired level of performance.

Balanced Scorecard (BSC) indicators

For BSC indicators, performance on each indicator is banded onto a scale of one to five - with 5 awarded for good performance and 1 for poor.

In 2004-05 three methods were applied to band the Balanced Scorecard indicators:

a. The absolute threshold method (as for Key Targets)

b. The mapping score method (for indicators with only a few value choices, e.g. 1, 2, 3)

c. The statistical confidence interval method whereby performance is assessed on whether there is statistical evidence that a trust's value is different from an expected range of values based on values from all trusts. This method was used for clinical, survey and change rate indicators.

All indicators are equally weighted within their focus area by simply summing up indicator band scores to produce the total score for the trust. These aggregated scores are then split into 3 bands (0, 1, and 2) according to specific thresholds for each Focus Area. In 2004-05 the approach to deriving these thresholds point was the same as that used in the previous year.

For the acute and specialist trusts, the overall scores were awarded as follows:

- trusts whose performance is in the top band (equal or above 45th percentile): two points
- trusts whose performance is in the middle band (17th-45th percentile): one point
- trusts in the bottom band (below 17th. percentile): zero points

These overall scores for each focus area: patient, clinical and capacity and capability are added together to give a score for the Balanced Scorecard on a scale of zero to six where a score of six indicates the highest level of performance on Balanced Scorecard indicators. This approach means that the contribution of each focus area is given equal weighting.

Source: Commission for Health Improvement/Healthcare Commission

APPENDIX 3 QUALITY OF SERVICES: NONFINANCIAL MEASURES IN THE ANNUAL HEALTH CHECK

	Quality of Services[55]		
	Existing national targets		
	2005-06	**2006-07**	**2007-08**
1	Total time in A& E: four hours or less	Total time in A& E: four hours or less	Total time in A& E: four hours or less
2	All cancers: 2 week wait	All cancers: 2 week wait	All cancers: 2 week wait
3	Outpatients waiting longer than the standard	Outpatients waiting longer than the standard	Outpatients waiting longer than the standard
4	Inpatients waiting longer than the standard	Inpatients waiting longer than the standard	Inpatients waiting longer than the standard
5	Cancelled operations	Cancelled operations	Cancelled operations
6	All cancers: 1 month diagnosis to treatment	All cancers: 1 month diagnosis to treatment	All cancers: 1 month diagnosis to treatment
7	All cancers: 2 month GP urgent referral to treatment	All cancers: 2 month GP urgent referral to treatment	All cancers: 2 month GP urgent referral to treatment
8	Revascularisation waits (3 months)	Revascularisation waits (3 months)	Revascularisation waits (3 months)
9	Thrombolysis: 60 minute call to needle time	Thrombolysis: 60 minute call to needle time	Thrombolysis: 60 minute call to needle time
10	Rapid access chest pain clinic waits	Rapid access chest pain clinic waits	Rapid access chest pain clinic waits
11	Convenience and choice: patient booking + information	Convenience and choice: patient booking + information	Convenience and choice: patient booking + information
12	Delayed transfers	Delayed transfers	

Continued …..

[55] Sourced from the Healthcare Commission's published performance ratings accessible at: http://www.cqc.org.uk/guidanceforprofessionals/nhstrusts/annualassessments/annualhealthcheck2005/06-2008/09.cfm

	APPENDIX 3 Quality of Services cont'd		
	Core standards[56]		
	2005-06	**2006-07**	**2007-08**
	Domain 1 Safety	***Domain 1 Safety***	***Domain 1 Safety***
	C1: Patient protection systems	C1: Patient protection systems	C1: Patient protection systems
	C2: Child protection compliance	C2: Child protection compliance	C2: Child protection compliance
	C3: NICE compliance	C3: NICE compliance	C3: NICE compliance
	C4: Hygiene and cleanliness	C4: Hygiene and cleanliness	C4: Hygiene and cleanliness
	Domain 2 Clinical and cost effectiveness	***Domain 2 Clinical and cost effectiveness***	***Domain 2 Clinical and cost effectiveness***
	C5: NICE technology appraisals	C5: NICE technology appraisals	C5: NICE technology appraisals
	C6: Working in partnership	C6: Working in partnership	C6: Working in partnership
	Domain 3 Governance	***Domain 3 Governance***	***Domain 3 Governance***
	C7: Clinical and corporate governance	C7: Clinical and corporate governance	C7: Clinical and corporate governance
	C8: Staff support	C8: Staff support	C8: Staff support
	C9: Records management	C9: Records management	C9: Records management
	C10: Employment checks	C10: Employment checks	C10: Employment checks
	C11: Recruitment, training and qualifications	C11: Recruitment, training and qualifications	C11: Recruitment, training and qualifications
	C12: Research governance	C12: Research governance	C12: Research governance
	Domain 4 Patient focus	***Domain 4 Patient focus***	***Domain 4 Patient focus***
	C13: Dignity and respect	C13: Dignity and respect	C13: Dignity and respect
	C14: Patient information on making a complaint	C14: Patient information on making a complaint	C14: Patient information on making a complaint
	C15: Hospital food	C15: Hospital food	C15: Hospital food
	C16: Patient information on services	C16: Patient information on services	C16: Patient information on services

[56] Many of the core standards had more than one subsidiary measure resulting in more than 40 measures in total.

	APPENDIX 3 Quality of Services – cont'd		
	Core standards – cont'd		
	2005-06	**2006-07**	**2007-08**
	Domain 5 Accessible and responsive care	***Domain 5 Accessible and responsive care***	***Domain 5 Accessible and responsive care***
	C17: Patient and carer consultation	C17: Patient and carer consultation	C17: Patient and carer consultation
	C18: Equality and equity	C18: Equality and equity	C18: Equality and equity
	C19: Emergency access	C19: Emergency access	C19: Emergency access
	Domain 6 Care environment and amenities	***Domain 6 Care environment and amenities***	***Domain 6 Care environment and amenities***
	C20: Safe and secure environment; privacy	C20: Safe and secure environment; privacy	C20: Safe and secure environment; privacy
	C21: Environment for effective care	C21: Environment for effective care	C21: Environment for effective care
	Domain 7 Public health	***Domain 7 Public health***	***Domain 7 Public health***
	C22: Partnership	C22: Partnership	C22: Partnership
	C23: Disease prevention and health promotion	C23: Disease prevention and health promotion	C23: Disease prevention and health promotion
	C24: Incident and emergency procedures	C24: Incident and emergency procedures	C24: Incident and emergency procedures
	New National targets		
1	Audit participation	Audit participation	Audit participation
2	Access to GUM clinics	Access to GUM clinics	Access to GUM clinics
3	Patient experience	Patient experience	Patient experience
4	Emergency bed days	Emergency bed days	Emergency bed days
5	MRSA	MRSA	MRSA
6	Ethnic group data quality	Ethnic group data quality	Ethnic group data quality
7	Self harm NICE compliance	Self harm NICE compliance	Self harm NICE compliance
8	Obesity: identification and management	Obesity: identification and management	Obesity: identification and management
9	Drug misusers: information, screening, referral	Drug misusers: information, screening, referral	Drug misusers: information, screening, referral
10	Smoke free NHS	Clostridium difficile data quality	Clostridium difficile data quality
11	Infant health: data completeness	Waiting times for diagnostic tests	Waiting times for diagnostic tests
12	MRI and CT scan waits	Referral to treatment time milestones	Referral to treatment time milestones
13		Smoking during pregnancy/breastfeeding	Smoking during pregnancy/breastfeeding

APPENDIX 4 USE OF RESOURCES MEASURES 2006-2008

1	**Financial reporting and external accountability**
1.1	Production of annual report
1.2	Availability and transparency of public reporting
2	**Financial Management**
2.1	Medium term financial strategy
	Medium term financial strategy
	Corporate and financial planning links
	Budget setting
	Cash management
	Savings plan and cost improvements
2.2	Performance against budgets
	Budget monitoring process
	Budget monitoring information
	Partnerships
	Planned savings
2.3	Management of asset base
	Estate strategy
	Asset register
	Asset management
3	**Financial Standing**
3.1	Financial breakeven
3.2	Financial projections
4	**Internal control**
4.1	Business risks
	Assurance framework
	Risk management
4.2	Internal control systems
	Statement of internal control
	Systems of internal control
	Audit Committee
	Internal audit
	Partnership arrangements
4.3	Probity and propriety
	Codes of conduct
	Counter fraud policy and culture
	Local counter fraud specialist

Continued…….

5	**Value for money**
5.1	Strategic and operational objectives
	Setting objectives
	Reviewing objectives
	Implementing objectives
5.2	Stakeholder needs
	Communication
	Patient feedback
	Engaging with hard to reach groups
5.3	Monitoring and reviewing performance
	Performance management process
	Data quality
5.4	Demonstration of value for money
	Efficiency plans
	Corporate and back office functions
	Clinical services
	Key NHS reforms
	Efficiency and productivity metrics
	Other benchmarking
	Reference costs/service line reporting
	Procurement
	SUMMARY: 5 perspectives; 14 main measures; 42 measures in total